An Evaluation of the Screening Approaches for Gestational Diabetes Mellitus

By

Qing Fang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Health Sciences

University of Warwick, Warwick Medical School
September 2016

Table of Contents

List of Figures

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 1

Chapter 3

Chapter 4

Chapter 5

List of Abbreviations and Acronyms

ACHOIS: Australian Carbohydrate Intolerance Study in Pregnant Women

ACOG: The American College of Obstetricians and Gynecologists

ADA: American Diabetes Association

BMI: body mass index

BDA: British Diabetes Association

CI: confidence interval

DM: diabetes mellitus

FPG: Fasting plasma glucose

GCT: Glucose Challenge Test

GDM: gestational diabetes mellitus

GWAS: genome-wide association analysis

HAPO: Hyperglycemia and Adverse Pregnancy Outcomes

HOMA IR: homeostatic model assessment for insulin resistance

HTA: health technology assessment

IADPSG: International Association of Diabetes and Pregnancy Study Group

ICER: incremental cost-effectiveness ratio

IGT: impaired glucose tolerance

IGTT: impaired glucose tolerance test

LGA: large for gestational age

N/A: not available

NDDG: National Diabetes Data Group (USA)

NHS: National Health Service (UK)

NICE: The National Institute for Health and Care Excellence (UK)

NPV: negative predictive value

NSC: National Screening Committee

OGTT: oral glucose tolerance test

OR: odds ratio

RCT: randomised controlled trial

SCI: Science Citation Index

SSCI: Social Sciences Citation Index

PCOS: polycystic ovary syndrome

PPV: positive predictive value

RCT: randomised controlled trial

ROC: receiver-operator characteristic

WHO: World Health Organization

Acknowledgment

I would like to thank a number of people whose support over the time it took to complete this research and prepare the thesis was of immeasurable value.

Firstly, I would like to express my deep and most sincere gratitude to my supervisor Dr Paul Sutcliffe for his detailed and constructive comments, encouragement and support throughout my PhD research. His enormous help and guidance during the thesis preparation have been invaluable. I am also deeply grateful to my co-supervisor Dr Saravana Ponnusamy for his extremely wide and insightful knowledge in gestational diabetes mellitus. His capacity for logical thinking benefitted me immensely. I must also acknowledge the help of my other co-supervisor Neil Raymond for his valuable advice on the study design and methodology for my research.

I wish to express my warmest thanks to the research collaborators from the Chengdu First People's Hospital in China, who provided me access to their patients and anonymised patents data for my primary research. Specifically, I would like to thank Yan Chen, Dr Xiaoxue Qi and Dr Xia Gu at the Department of Gynaecology and Obstetrics for their wonderful support.

My sincere appreciation also goes to Dr Martin Connock, Dr Peter Kimani and Riwa Meshaka at the University of Warwick for their helpful advice on the analysis of the systematic review, the case-control study and the Q methodology study. I would also like to acknowledge the help of Professor Jeremy Wyatt at the NIHR Evaluation, Trials and Studies Coordinating Centre for his generous advice on conducting the diagnostic test accuracy study analysis.

Finally, I would like to thank my parents for their immense love and support throughout my PhD and particularly during the preparation of this thesis. Indeed, their encouragement and support made this journey less arduous than it would have been otherwise.

Declaration

I declare that this thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy by the High Degree Committee at the University of Warwick. The thesis has been composed and written by myself based on my own work. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given. This thesis has not been submitted in any previous application for a higher degree.

Qing Fang
Division of Health Sciences
Warwick Medical School
University of Warwick
England, UK

Abstract

Background: Gestational diabetes mellitus (GDM) is defined as any degree of glucose intolerance that occurs or is first recognised during pregnancy. The prevalence of GDM is 1-28% globally and 11% in China. Although GDM can cause severe maternal and neonatal outcomes, there is no consensus worldwide as to whether universal or selective screening of expectant mothers should be recommended. In 2010, The International Association of Diabetes and Pregnancy Study Groups (IADPSG) recommended that all pregnant women should be screened via a one-step universal screening approach for GDM, using a 75g oral glucose tolerance test (OGTT) with reduced thresholds. Despite ongoing debate over the efficacy and use of the IADPSG approach, China was the first country to adopt the new screening approach. A number of observational studies have shown that the new IADSPG approach is clinically more effective. However, reservations exist as to the associated increase in health costs and inconvenience to pregnant women.

Aim: To assess and explore the best screening approach for GDM both globally and in China.

Methods: The research involved three projects. Project I (Chapter 3) was a systematic review of the effectiveness and cost-effectiveness of universal versus selective screening for GDM, which followed a standard systematic review procedure for Diagnostic Test Accuracy studies. Project II (Chapter 4) was a Q methodology study to investigate the pregnant women's attitudes towards and experience of the IADPSG one-step screening approach for GDM in China. A total of 30 pregnant women who visited the hospital for antenatal care in 2014 were recruited to participate in the study. The Q methodology study was undertaken using the FlashQ software and were analysed using the PQMethod software. Project III (Chapter 5) was a case-control study to establish and assess a risk score algorithm in order to improve the IADPSG approach for GDM screening in China. Medical records of 550 pregnant women (272 GDM cases and 278 controls) who had given birth in the year 2013 at the Chengdu First People's Hospital were retrospectively collected and analysed. Univariable analysis and multiple logistic regression analysis were used to identify GDM risk factors and to formulate the risk score algorithm. A

Receiver Operating Characteristic (ROC) curve was employed to assess the effectiveness of the risk score algorithm for GDM screening.

Results: The systematic review (Chapter 3) included 28 effectiveness studies, four cost studies and one cost-effectiveness study. Seven out of the 28 effectiveness studies and the cost-effectiveness study favoured selective screening. The Q methodology study (Chapter 4) suggested that the participants agreed as to the importance and necessity of the IADPSG one-step GDM screening for all pregnant women. However, the non-GDM women felt somewhat burdened in undertaking the fasting and 2-hour oral Glucose Tolerance Test (OGTT) for GDM under the IADPSG approach. The participants also desired more information on GDM and OGTT both before and after the test. The case-control study (Chapter 5) identified age, height, body mass index (BMI), family history of diabetes, waist circumference, previous deliveries and blood pressure before 24^{th} week of gestation to be risk factors for GDM in the Chinese population. Subsequently, a risk score algorithm was established, whereby the use of the risk score to select high-risk women for screening could help to exclude nearly half (45%) of non-GDM women from the OGTT while still diagnosing 80% of the GDM cases.

Conclusion: Universal screening for GDM is recommended for areas where GDM prevalence is relatively high and where economic constraints circumscribing implementation of the approach do not exist. For areas where GDM prevalence is low, it is recommended that current practice, whether it is universal or selective screening, should be retained until more robust evidence emerges. The IADPSG one-step universal screening was viewed positively in terms of importance and necessity by participants of the study, and they felt that GDM screening is necessary to be undergone by every pregnant woman. At the same time, the non-GDM women also felt strongly that the two-hour OGTT requiring 3 blood samples over the test period was inconvenient and burdensome. Alternatively, the use of a risk score-based selective IADPSG approach was observed to be conducive to the exemption of nearly half (45%) of non-GDM women from the OGTT test while still diagnosing 80% of the GDM cases in China. A future validation cohort from other parts of China is required to affirm the effectiveness of this risk scoring algorithm.

Executive Summary

GDM is the onset or first recognition of glucose intolerance during pregnancy. GDM can lead to severe maternal and neonatal outcomes including pre-eclampsia, caesarean section delivery, shoulder dystocia, macrosomia and stillbirth. GDM screening can facilitate early treatment and reduce negative outcomes. However, whether all pregnant women should be screened (universal screening) or only high-risk women should be screened (selective screening) for GDM is a widely debated issue. In 2010, the IADPSG suggested a new universal screening approach requiring every pregnant woman to undergo the 75g Oral Glucose Tolerance Test (OGTT) for GDM diagnosis between the 24^{th} to 28^{th} week of gestation. The thresholds of the OGTT were reduced to diagnose more pregnant women with GDM. China was the first country to adopt the IADPSG approach, despite the fact that clear consensus does not exist as to the best screening approach or adoption of the new IADPSG one-step approach for GDM. The PhD research assessed the effectiveness and cost-effectiveness of universal versus selective screening, explored patient perspectives on IADPSG and tested a risk score-based selective screening approach under the new IADPSG criteria.

Chapter 1: This chapter provides a background to the PhD research, introducing GDM as a growing health problem and the role of GDM screening in addressing this issue. It describes different GDM screening approaches, the controversies in the evidence and implementation across the globe. It highlights the challenges in addressing the inconsistency pertaining to the identification of the optimal GDM screening approach, identifies the research gaps and justifies the rationale for this PhD research.

Chapter 2: This chapter outlines the overall research question and the aims and objectives of this research.

Chapter 3: This chapter presents a systematic review of the effectiveness and cost-effectiveness of selective screening in comparison with universal screening for GDM. Four electronic databases were searched, including Medline, Embase, Science

Citation Index (SCI) and Social Sciences Citation Index (SSCI) (via Web of Science) and the Cochrane database. For effectiveness studies, the outcome measures selected were specificity and sensitivity of selective screening in comparison with universal screening. Thirty three studies were included in the review, including 28 effectiveness studies, four cost studies and a single cost-effectiveness study. Only seven of the 28 effectiveness studies recommended selective screening for pregnant women. These seven studies were conducted in areas with relatively low GDM prevalence and included the four studies using their own selection criteria rather than standard guidelines for identifying high risk women. The cost-effectiveness study and three of the four cost studies found that universal screening was slightly more expensive and less cost-effective when compared to selective screening. In conclusion, universal screening was recommended for areas with relatively high GDM prevalence and absence of economic constraints. In areas where GDM prevalence is low, it is recommended that current practices, whether led by universal or selective screening, are maintained until stronger evidence in support of either emerges. This chapter reports that selective screening with self-developed selection criteria for high risk women based on local population could potentially be effective and recommends future research in this direction. It discusses the limited evidence as to the cost-effectiveness of selective screening in comparison with universal screening.

Chapter 4: This chapter presents the findings of a Q methodology study focusing on pregnant women's attitudes, views and experience of the IADPSG universal approach for GDM screening in China. In 2011, China adopted the IADPSG recommendation and the one-step universal approach, thereby replacing the previous two-step universal approach that comprised a screening test followed by OGTT if needed. Non-GDM women who would have only received the simple screening test (i.e., 50g glucose challenge test) under the two-step approach now had to undertake the complicated OGTT test. Also, the number of women diagnosed with GDM was estimated to double or triple under the new IADPSG approach. Q methodology is used to explore subjective perspectives including participant points of view, opinions, beliefs and attitudes. The Q methodology study was conducted to obtain the perspectives of pregnant women about the new IADPSG approach in order to

improve the approach in accordance with the needs of expectant mothers. A total of 32 Q statements (Q-set) relevant to attitudes, views and experience of the GDM testing approach were developed from a review of literature, consultations with pregnant women, expert opinions and online sources. The FlashQ software was used for computerising the Q statements to facilitate data collection and to provide a user friendly interface. Thirty pregnant women (15 GDM and 15 non-GDM women) from the Chengdu First People's Hospital were recruited (P-set) to participate in the study. They were asked to rank-order the 32 Q statements from agree least to agree most (-3~+3) based on their own viewpoints into a table of normal distribution (Q sorts) and provide reasons for the statements they most agreed and disagreed with. The PQMethod software was used for Q sorts analysis, and the reasons they provided were used for interpretation and illustration. Two distinct shared viewpoints emerged from the Q analysis results. In general, pregnant women tended to agree that GDM screening was important and necessary and needed to be administered to all pregnant women. However, the non-GDM women felt strongly that OGTT was inconvenient in terms of the three blood samples required and prolonged duration of two hours needed to complete the test, thereby leaving them to feel burdened in undergoing the OGTT. Both GDM and non-GDM women wished strongly to be provided with more information on GDM and OGTT both before and after undergoing the OGTT.

Chapter 5: This chapter presents a nested case-control study that established and assessed the effectiveness of a risk score-based selective screening approach for GDM led by IADPSG criteria in China. Studies showed that while the IADPSG approach was clinically more effective, there were global concerns about the increased costs and burden associated with approach. This study explored the use of existing and novel risk factors to generate a risk scoring algorithm to select only the high-risk pregnant women for GDM screening. Medical records of 2897 pregnant women who gave birth during 2013 while receiving care at the Chengdu First People's Hospital were investigated. The study found that GDM prevalence was 9.4%. A nested case-control study was conducted involving 272 GDM and 278 non-GDM. It was identified that age, height, body mass index (BMI), family history of diabetes, waist circumference, previous deliveries, systolic and diastolic blood pressure during the first trimester represented risk factors for GDM in the Chinese

population. A risk scoring algorithm was formulated using the adjusted odds ratios of these eight risk factors from the multiple logistic regression result. A Receiver Operating Characteristic (ROC) curve was drawn to assess the effectiveness of the algorithm, showing that a cut-off score of 0.32 provided an optimal sensitivity of 80% and a specificity of 45%. This implied that risk score-based selective screening would correctly identify 80% of the GDM women as high-risk women through the use of the risk scoring algorithm, enabling the mothers-to-be to undergo the OGTT. On the other hand, 20% GDM women (2% of the pregnant population) would be missed due to this kind of screening. Essentially, 45% of the non-GDM women (41% of the pregnant population) would be correctly identified as low risk women, thereby avoiding the OGTT. A further cohort from other areas of China is needed to affirm the effectiveness of the risk scoring algorithm.

Chapter 6: This final chapter summarises the key findings of this research and discusses their implications for practice and for future research. The optimal GDM screening approach is "setting dependent". It is influenced by the population characteristics and risks, the different screening tests and criteria used and the patient perspectives as well as the wider context of healthcare value and culture. Until more evidence emerges, universal screening is recommended for areas where GDM prevalence is relatively high and without economic constraints. Whether it is universal or selective screening, current practice should be maintained for areas where GDM prevalence is low. In China, the new IADPSG universal screening approach was generally accepted by pregnant women, although non-GDM women largely felt that the IADPSG screening test was inconvenient and burdensome. The exploration of a risk score-based selective screening approach under the IADPSG criteria showed that nearly half of pregnant women would be exempted from the GDM screening test, while 80% of the GDM cases would be identified accurately in China. Future research is also needed to provide more evidence as to the cost-effectiveness of selective versus universal screening. Evaluations on the effectiveness of a risk-scoring algorithm and patient perspectives for GDM screening within country-specific settings are strongly recommended.

Chapter 1: Introduction

1.1 INTRODUCTION

'There is an urgent need for universally applicable simple screening and diagnostic procedures criteria for GDM – lessons from projects funded by the World Diabetes Foundation'.

(Nielsen *et al.*, 2012)

This quote reflects the global debate as well as the complexity undergirding the optimal screening and diagnostic approaches for gestational diabetes mellitus (GDM). In the absence of international consensus, different guidelines and implementations for GDM screening exist between and even within countries. In 2010, the International Association of Diabetes and Pregnancy Study Groups (IADPSG) recommended a universal screening approach for GDM with reduced test thresholds, which caused even more debate and controversy worldwide. Nevertheless, China was the first and one of the few countries which adopted the IADSPG criteria. The evaluation of the effectiveness, cost-effectiveness and user-perspectives of the IADPSG approach is at its very initial stage. As a Chinese citizen and health researcher, I developed a strong interest in this particular area as I was keen to evaluate different GDM screening approaches so as to address the inconsistency and contribute to the evidence-based recommendations in GDM screening approaches.

At present, there are worldwide uncertainties about: 1) whether all pregnant women should be screened (universal screening) or only screen women with GDM risk factors (selective screening); 2) should the new IADPSG one-step screening approach for GDM be adopted; and 3) especially for China, what implications there are for adopting the IADPSG approach and what can be improved. This introduction summarises and critically appraises the recent debate on and implementation of GDM screening approaches and identifies gaps of research in this area.

1.2 GDM AS A HEALTH PROBLEM

1.2.1 Definition of GDM and adverse outcomes of the condition

GDM was generally understood to represent "any degree of glucose intolerance that occurs or is first recognised during pregnancy" (Metzger & Coustan, 1998). Contemporary definitions have evolved to refer to GDM as diabetes diagnosed during pregnancy that is not clearly overt diabetes (American Diabetes Association, 2013). Generally asymptomatic, GDM can nevertheless cause adverse short and long-term maternal and fetal outcomes (Langer *et al.*, 2005). The Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study serves as a pivotal clinical study for understanding the existence of a continuous relationship between mild hyperglycaemia and adverse pregnancy outcomes (Test, 2008).

Adverse consequences to maternal health include spontaneous abortion, gestational hypertension, pre-eclampsia and caesarean section, while stillbirth, macrosomia, large for gestational age (LGA) babies, birth trauma and shoulder dystocia may be any one of the conditions experienced by the babies of mothers with GDM (Ju *et al.*, 2008; Langer *et al.*, 2005; Odar *et al.*, 2004; Test, 2008; Reece *et al.*, 2009). Additionally, women with GDM and their offspring are at increased risk of developing type 2 diabetes in the future (Damm *et al.*, 2009; Kim *et al.*, 2002).

1.2.2 Aetiology and pathology of GDM

Pregnancy is a condition characterised by progressive insulin resistance that commences close to the midpoint of the gestational period and progressed through the third trimester. Insulin sensitivity falls by up to 50% in late pregnancy (Di Cianni

et al., 2003). It has been noted that the main contributors to insulin resistance include the insulin desensitising effects of hormones produced by the placenta, followed by increased maternal adiposity (Perkins *et al.*, 2007). During normal pregnancy, resistance to insulin action is observed to increase, as in most pregnancies, pancreatic beta cells are able to compensate for increased insulin demands, thereby maintaining normoglycaemia. In contrast, women who develop GDM experience deficits in beta-cell response, thereby leading to insufficient insulin secretion that can compensate for increased insulin demands. The risk for GDM is increased by many factors including: age, ethnicity, obesity, first-degree relatives with diabetes, previous gestational diabetes, gestational weight gain, polycystic ovarian syndrome (PCOS), smoking, and many others (Association, 1997; Force, 2008; Walker, 2008; Webber *et al*., 2015; NICE 2015). The high risk ethnic origins include South Asia (India, Pakistan, Bangladesh), Black Caribbean and Middle Eastern origin (Saudi Arabia, United Arab Emirates, Iraq, Jordan, Syria, Oman, Qatar, Kuwait, Lebanon, Egypt) (Webber *et al*., 2015).

1.2.3 Epidemiology of GDM

With the increasing global epidemic of obesity and type 2 diabetes, as well as the increasing age of conception and child bearing, the prevalence of GDM is rising (Aljohani *et al.*, 2008; Anna *et al.*, 2008; Ferrara, 2007; Waugh *et al*., 2010). However, estimating the GDM prevalence is made difficult by a lack of universally accepted diagnostic criteria for GDM, a factor which also hinders the consistent diagnosis and management of GDM in clinical practice (Reece *et al.*, 2009). GDM incidence is expected to further increase with the adoption of the new screening criteria proposed by the IADPSG (Panel, 2010).

GDM affects 1–28% pregnant women worldwide, with wide variations due to ethnicity (Jiwani *et al.*, 2012). In Europe, GDM affects 0.15–4% of pregnant women (Baliutaviciene *et al.*, 2002). In the UK, up to 5% of the women giving birth in England and Wales were reported to have diabetes, with 87.5% of the expectant mothers within this percentage having GDM (NICE, 2015). In China, the GDM incidence was reported to range from 8% to 15% calculated according to the new IADPSG approach for GDM diagnosis (Shang & Ma, 2011; Wei & Yang, 2011; Hou *et al.*, 2012; Lu *et al.*, 2012; Jiang *et al.*, 2013).

1.2.4 Screening for GDM

GDM is detected through screening carried out on pregnant women between the 24th to 28th gestational week. In 2002, Scott *et al.* (2002) examined the evidence of GDM screening against the 10 criteria identified by the UK National Screening Committee (NSC) for determining whether a screening programme should be implemented, and they found that GDM screening did not meet all the NSC criteria. Before 2007, the benefit of screening pregnant women for GDM was not clear and not adequately supported by rigorous scientific evidence (Vidaeff *et al.*, 2003; Russell *et al.*, 2007). Studies and reviews carried out subsequently provided increasing evidence in support of benefits associated with GDM screening. In 2008, a cost-effectiveness study by the UK National Institute for Health and Care Excellence (NICE) concluded that it was cost-effective to conduct GDM screening (NICE, 2008). This study used data from an earlier Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) carried out by Crowther *et al.* (2005). Recently, the U.S. Preventive Services Task Force (USPSTF, 2013) found that there was a moderate net benefit associated with screening for GDM after 24th week of gestation so as to reduce maternal and fetal complications, including the collective outcomes of preeclampsia, macrosomia, and shoulder dystocia. Hence, in light of the support

available in literature, it can be concluded that GDM screening is necessary for preventing the development of adverse maternal and neonatal outcomes.

The two alternative test(s) for GDM screening currently available are the two-step tests or a one-step test. For two-step tests, pregnant women first undergo a screening test of 50g glucose challenge test (GCT), women with a positive test result further undergo a diagnostic test, a 75g or 100g oral glucose tolerance test (OGTT). For the one-step test, pregnant women undergo the OGTT directly for diagnosis. The 50g GCT is the most commonly used screening test, wherein the venous plasma glucose is tested one hour after drinking 50g of glucose in solution. Other alternative screening tests include glycosylated hemoglobin (HbA1c) and fasting plasma glucose (FPG) (Hartling *et al.*, 2012). The diagnostic test is 75g or 100g OGTT, within which the venous plasma glucose is tested at fasting state, and then 1 hour, 2 hours, or 3 hours after drinking 75g or 100g of glucose in solution. There are several different OGTT criteria prevalent internationally, which are summarised in Table 1.

Table 1. Different diagnostic criteria for GDM screening internationally

	Diagnostic test	**Venous plasma mmol/l (mg/dl)**			
		fasting	**1h**	**2h**	**3h**
IADPSG (Panel, 2010)	75g OGTT	5.1 (92)	10.0 (180)	8.5 (153)	
World Health Organization (Organization, 1999)	75g OGTT	7.0 (126)		7.8 (140)	
O'Sullivan and Mahan (O'Sullivan & Mahan, 1964)	100g OGTT	5.0 (90)	9.1 (164)	8.0 (144)	6.9 (124)
National Diabetes Data Group (Group, 1979)	100g OGTT	5.8 (105)	10.6 (190)	9.2 (165)	8.1 (145)
Carpenter and Coustan/ American Diabetes Association (Association, 2010)	100g OGTT	5.3 (95)	10.0 (180)	8.6 (155)	7.8 (140)

1.2.5 Treatment and management of GDM

Studies have shown that the risk of adverse pregnancy outcomes can be prevented or reduced by achieving glycaemic control alongwith lifestyle modifications and/or pharmaceutical intervention during pregnancy (Crowther *et al*., 2005; Landon *et al*., 2009). Such lifestyle modifications typically involve dietary changes and exercise, whereas pharmacological treatment is reserved for women who are unable to maintain an acceptable range of glucose levels despite adjustments in terms of diet and exercise (Chirayath, 2006; Holt *et al*, 2008).

The treatment of GDM has been observed to be effective in reducing macrosomia, preeclampsia and shoulder dystocia (Falavigna *et al.*, 2012). A multicenter randomised trial conducted in the US demonstrated that it was potentially useful to treat pregnant women with mild GDM (Landon *et al.*, 2009). They found that though treatment of mild GDM did not significantly reduce the frequency of stillbirth or perinatal death and other neonatal complications, it did reduce the risks of fetal overgrowth, shoulder systocia, cesarean delivery and hypertensive disorders.

Further, the Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) trial showed reductions in perinatal complications among infants born to mothers who were provided with more intensive dietary advice, blood glucose monitoring and insulin upon requirement (Crowther *et al.*, 2005). In connection with the treatment for GDM, the HAPO study showed that women failing to control hyperglycaemia in pregnancy by lifestyle measures alone could be safely and effectively treated with oral agents, for instance metformin or glibenclamide, rather than being directly administered insulin (Test, 2008).

Waugh *et al.* (2010) synthesised evidence pertaining to the risks and benefits of oral glucose lowering drugs in comparison with insulin in the treatment of hyperglycaemia in pregnancy. The RCT evidence showed little difference in results between the drugs and insulin. When comparing metformin with insulin, less maternal weight gain was reported with the use of metformin, but insulin evidenced better outcomes in terms of age at delivery. When comparing glibenclamide with insulin, there was evidence of less maternal hypoglycemia with glibenclamide, but insulin was observed to achieve less neonatal hypoglycaemia and lower birthweight. The review suggested that both metformin and glibenclamide were effective alternatives to insulin. Not surprisingly, some evidence suggested that women preferred oral agents to insulin treatment (Waugh *et al.*, 2010).

1.2.6 Postnatal care and follow-up of GDM

Pregnant women with GDM are unlikely to require insulin after the delivery, so all hypoglycemic agents (either oral or insulin) should be stopped immediately after birth. With the purpose of detecting women with previously undiagnosed pre-existing diabetes, it is suggested that pregnant women with GDM undergo blood glucose monitoring in the early postnatal period (Luesley & Mark, 2016).

Several clinical trials have concluded that lifestyle modifications and/or pharmacological intervention have the potential to prevent pregnant women with a history of GDM from progressing to type 2 diabetes (Ratner *et al.*, 2008; Xiang *et al.*, 2006). Postnatal follow-up of GDM and corresponding necessary preventative interventions have been of importance to preventing long-term health consequences for both the mother and her child.

1.3 GDM SCREENING APPROACHES

1.3.1 Controversies over GDM screening approaches

There is continued disagreement and inconsistency regarding the best screening strategy (Scott *et al.*, 2002; Waugh *et al.*, 2010). The current evidence is inconclusive to make a recommendation on whether universal or selective screening should be implemented (Tieu *et al.*, 2010). Universal screening for GDM was recommended by the International Workshop Conference on Gestational Diabetes Mellitus in 1980, 1984, and 1990 (Baliutaviciene *et al.*, 2002). While in 1997, the fourth International Workshop Conference on Gestational Diabetes Mellitus and the American Diabetes Association (ADA) recommended selective screening for GDM (ADA, 1997). The fifth international workshop conference in 2007 continued to recommend selective screening for GDM (Metzger *et al.*, 2007). However, the ADA changed its previous recommendation from selective screening (Association, 2010) to universal screening in 2011 (Prevention & TYPE, 2011). Meanwhile, as reported in Section 1.2.4, either two-step tests or a one-step test is used for GDM screening. Figure 1 illustrates the different screening approaches and tests used.

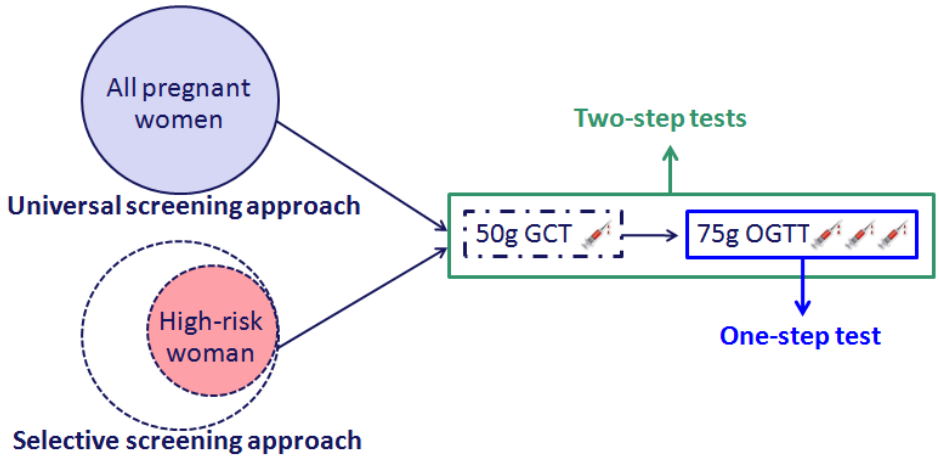


Figure 1. Different GDM screening approaches and testing methods

Under the selective screening approach, women are categorised into high risk and low risk individuals. Subsequently, only the high risk women undergo GDM screening test under the selective approach. High risk women are identified on the basis of their risk factors for developing GDM. Table 2 shows the commonly used selection criteria according to the NICE and ADA guidelines, which are also mentioned in section 3.1.1. If a pregnant woman presents one of these risk factors, she is classified as a high risk woman and undergoes the screening test.

Table 2. Selection criteria for high risk women for GDM screening

GDM risk factors	ADA (2010)	NICE (2015)
Maternal age	>25 yrs	
Overweight (BMI)	√	>30
Family history of diabetes	√	√
History of GDM	√	√
History of macrosomia	√	≥4.5kg
Certain ethnic groups	√	√
Adverse pregnancy history	√	
Others	Glucosuria, PCOS	

In 2010, the International Association of Diabetes in Pregnancy Study Group (IADPSG) recommended a one-step universal screening approach with reduced OGTT thresholds for GDM. This recommended a 75g OGTT at 24–28 weeks for all women not previously diagnosed with diabetes by random or fasting plasma glucose testing at the first antenatal visit (Panel, 2010). The thresholds of 75g OGTT were reduced to diagnose more GDM women who might develop adverse maternal and neonatal outcomes. The lowered thresholds were ≥5.1 mmol/l at fasting, 10 mmol/l at 1 hour and 8.5 mmol/l at 2 hours. A diagnosis of GDM was made if at least one abnormal value was identified. Before the IADSPG recommendation, the commonly used criterion was that GDM diagnosis was confirmed with at least two abnormal

values from four measurements (blood glucose ≥5.8mmol/l at fasting, 10.6mmol/l at 1 hour, 9.2mmol/l at 2 hours, and 8.1 mmol/l at 3 hours).

Though the one-step universal approach for GDM was suggested by the IADPSG, there were many concerns related to implementing the approach. Vandorsten *et al.* (2012) have indicated that there is no clear evidence to demonstrate improvements in the health and patient-centered outcomes through the adoption of the one-step approach. They have voiced reservations about the increased patients costs, life disruptions and psychological burdens result from increase in the number of women diagnosed with GDM under the new approach. Langer *et al.* (2013) have pointed out that the IADPSG criteria has not been analysed systematically for medical, social, and economic ramifications, and caution should be exercised in promoting the new approach until associated outcomes have been rigorously evaluated.

The American Diabetes Association (ADA) adopted the IADPSG recommendation in 2010 (Association, 2010) and then rejected it in December 2013 (Association, 2013). The National Institutes of Health (NIH) Consensus Group rejected IADPSG in March 2013 (NIH, 2013). However, the World Health Organization (WHO) endorsed IADPSG in August 2013 (WHO, 2013). Worldwide debate continues over whether or not to adopt the IADPSG approach for GDM. In the context of China, the IADPSG one-step universal approach was adopted in July 2011 (Chinese Ministry of Health, 2011).

1.3.2 Different implementations among countries

Due to the lack of evidence and consensus on the IADPSG approach and the optimal approach for GDM screening, countries are implementing different screening practices (Figure 2). In the United States, universal screening with two-step tests is

commonly used with criteria proposed by the National Diabetes Data Group (NDDG) (Group, 1979) or Carpenter and Coustan (Carpenter & Coustan, 1982). Many developing countries are also implementing the universal screening approach, including Brazil, India, Mexico, and Ethiopia (Moyer, 2014; Tsatsoulis *et al.*, 2009). In the United Kingdom, there are variations in practice among different areas and hospitals, however, most of the hospitals conduct selective screening with a one-step 75g OGTT according to the NICE guideline (NICE, 2015). France and New Zealand are implementing selective screening with two-step tests (Vambergue, 2010; Simmons *et al.*, 2009). China has adopted the new IADPSG one-step universal approach since 2011 (Chinese Ministry of Health, 2011), and the new approach was implemented by the majority of hospitals in China. Previously, China recommended universal screening with two-step tests (Le, 2008). Italy adopted the IADPSG criteria in 2011 and Australia adopted it in 2014 (Wong, 2014).

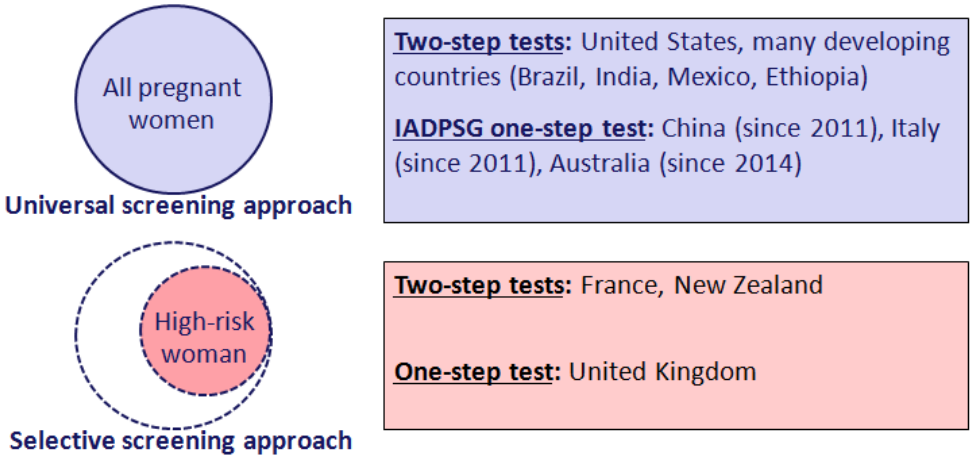


Figure 2. Global implementation of the different GDM screening approaches

1.3.3 Evidence on selective and universal screening approaches

It is still internationally unclear whether to conduct a universal or selective screening approach for GDM, or whether to adopt the IADPSG one-step universal approach.

The following sections aim to summarise the literature addressing these issues, and to present the current state of development in these areas.

Several reviews have been published, which partially addressed the issue of universal versus selective screening for GDM. In the review by Scott *et al.* (2002) for assessing GDM screening methods and costs, it identified two studies with direct comparison between routine universal screening and selective screening (Bebbington *et al.*, 1999; Casey *et al.*, 1997) and several studies of subgroups of women at higher risk of GDM, with a view to having selective rather than universal screening. It found that low sensitivities and specificities had been reported using risk factors alone as a screening test.

One recent systematic review (Tieu *et al.*, 2010) evaluated the screening, diagnosis and treatment of GDM, addressing the issue of universal versus selective screening. The review included only randomised and quasi-randomised trials, and identified one quasi-experimental study (Griffin *et al.*, 2000) which suggested universal screening was superior to selective screening and resulted in higher detection rate, earlier diagnosis, and better pregnancy outcomes. However, due to the risks of bias in the trial, the systematic review did not reach a definite conclusion about which screening approach should be recommended. As pointed out by Tieu *et al.* (2010), the trial had high risk in randomisation, allocation concealment and outcome data report, outcome data were analysed by GDM diagnosis rather than by original group allocation, affecting interpretation of outcome data. Another recent systematic review (Hieronimus & Le Meaux, 2010) also assessed studies relevant to selective versus universal screening. It concluded that the benefits of GDM screening and treatment had only been proven for women with GDM risk factors, their relevance in women without risk factors remained controversial because of the number of unnecessary tests and the cost-effectiveness ratio. However, the review (Hieronimus

& Le Meaux, 2010) only searched Medline and Cochrane Databases and did not conduct quality assessment for the studies included therein. Further information about systematic reviews in this area will be discussed in chapter 3.

1.3.4 Evidence on the new IADPSG universal screening approach

1.3.4.1 International studies on the clinical and cost implications of the new IADPSG approach

In 2008, the HAPO study investigated whether maternal hyperglycaemia less severe than that in diabetes mellitus is associated with increased risks of adverse pregnancy outcomes (Test, 2008). A total of 25,505 pregnant women at 15 centers in nine countries were involved. The HAPO study found strong, continuous associations of maternal glucose levels below the diagnosis of diabetes with increased birth weight and increased cord-blood serum C-peptide levels. This indicated that adverse maternal and neonatal outcomes appeared across a wider range of maternal glucose levels than previously thought. Having reviewed the result from the HAPO study, in 2010, the IADPSG Consensus Panel recommended a major change in GDM screening, promoting a one-step 75g OGTT for all pregnant women from 24^{th} to 28^{th} week of gestation with lower OGTT threshold values than had been used previously (Panel, 2010). Under the new regimen of one-step universal screening approach, GDM diagnosis is confirmed if there is at least one abnormal value registered from the three that have been measured (fasting glucose ≥5.1 mmol/l, 10 mmol/l at 1 hour, and 8.5 mmol/l at 2 hours). In line with this, the new IADPSG approach was expected to substantially increase the GDM incidence, potentially doubling or tripling the incidence (Panel, 2010).

One issue with the IADPSG criteria was that it is only based on observational datasets, including the large observational HAPO study (Test, 2008). No randomised controlled trial has been conducted to test the effectiveness of the IADPSG approach, which prevents a definitive conclusion of the superiority of IADPSG (Cundy *et al.*, 2014). However, following the observational study of HAPO (Test, 2008), there have been increasingly more observational studies (i.e., case-control studies) undertaken supporting the notion that the IADPSG is clinically more effective (Lapolla *et al.*, 2011; Benhalima *et al.*, 2013). These retrospective studies have found that the women classified as normal in accordance with the old criteria but re-classified as GDM by the IADPSG criteria (i.e., new GDM women) had significantly higher incidences in clinically important adverse outcomes (e.g., caesarean section delivery, large for gestational age, shoulder dystocia). This implied that diagnosing and treating these women would have improved the outcomes. In China, the conducted observational studies also showed the IADPSG approach was more clinically effective (Shang & Ma, 2011; Wei & Yang, 2011; Lu *et al.*, 2012; Jiang *et al.*, 2013).

Two recent studies evaluated the cost-effectiveness of the IADPSG one-step approach compared with the two-step approach currently used in many countries (Mission *et al.*, 2012; Werner *et al.*, 2012). Mission *et al.* (2012) found that the IADPSG approach was more expensive and more effective but cost-effective at $61,503/ quality-adjusted life year (QALY). However, Werner *et al.* (2012) suggested the IADPSG recommendation was cost-effective only when post-delivery care reduced diabetes incidence. Thus far, no cost-effectiveness study has been conducted in China. Vandorsten *et al.* (2012) have concluded that available studies do not provide clear evidence that a one-step approach is more cost-effective compared with the current two-step approach.

Concerns over the costs and cost-effectiveness are among the key reasons that countries hesitate to adopt the IADPSG approach. The NIH panel stated that they were particularly concerned about the increase of corresponding costs and interventions by adopting the IADPSG approach (NIH, 2013). The increased costs are comprised of the increased diagnosis costs by conducting OGTT among all pregnant women, and the increased healthcare costs by treating the additional women diagnosed as GDM. If, as concluded by all the observational studies, the IADPSG approach is more clinically effective, it will be necessary to spend the additional treatment expenditure on these GDMs. However, there is still a need to explore whether the diagnosis costs of conducting the OGTT could be reduced without compromising the clinical effectiveness, which would improve the cost-effectiveness of the IADPSG approach.

1.3.4.2 Chinese studies on the clinical and cost implications of the new IADPSG approach

The new IADPSG recommendation was based on the HAPO study that did not include data from China. To explore the applicability of the approach to China, several Chinese studies investigated the clinical effectiveness of the new IADPSG approach in comparison with other approaches. These are summarised below.

Three studies (Shang & Ma, 2011; Lu *et al*., 2012; Cai & Yang, 2012) compared the IADPSG one-step universal approach with the previous two-step universal approach for GDM. Shang & Ma (2011) and Lu *et al.* (2012) evaluated the clinical outcomes of the "over-diagnosed" GDM cases that were picked up by the IADPSG criteria but not by the previous two-step tests. Shang & Ma (2011) divided the additionally diagnosed GDM cases into treatment and non-treatment group, and found the treatment group had better clinical outcomes, indicating it was important to diagnose

and treat these additional GDM women. Lu *et al.* (2012) compared the outcomes of the "over-diagnosed" cases with non-GDM women and found that, if left untreated, the additionally diagnosed GDM cases were characterised by significantly higher rates of maternal and neonatal complications than women without GDM. Both findings suggested that IADPSG approach was clinically more effective in China. The other study by Cai & Yang (2012) was biased in study design. It found that if treatment were provided, the GDM diagnosed against the IADPSG criteria had significantly fewer adverse outcomes than GDM women diagnosed with the older criteria, thus leading to the conclusion that IADPSG criteria was better. However, this might simply have transpired because the GDM cases diagnosed by the IADPSG criteria were milder cases due to reduced OGTT thresholds.

Jiang *et al*. (2013) compared the one-step 75g OGTT universal approach using the IADPSG cut-offs (reduced thresholds) and the older cut-offs. The additional women who were diagnosed by the IADPSG cut-offs (8.14% GDM incidence) but not by the older cut-offs (4.57% GDM incidence) were divided into treatment and non-treatment groups. It was shown that the maternal and neonatal outcomes were significantly better in the treatment group, thereby suggesting that the reduced cut-offs of the IADPSG criteria improved clinical outcomes. However, no published cost or cost-effectiveness study for the IADPSG screening approach in China has been identified.

1.3.5 User perspectives on GDM screening and diagnosis

Patient-centred healthcare is increasingly emphasised in current research and guidelines (Bauman *et al*., 2003; Stiggelbout *et al*., 2012). Therefore, it is essential to understand the perspectives of pregnant women in GDM screening. As GDM screening involves pregnant women who are a special and vulnerable user group,

evidence from user perspectives and experience plays an important role in the design, implementation, and evaluation of a quality healthcare service. Patient attitudes, views, and experiences of GDM screening need to be considered and explored in the process of identifying and implementing the optimal GDM screening and diagnostic approach.

One Australian study has examined women's attitudes towards a universal one-step GDM screening with modified OGTT (Griffiths *et al.*, 1993). The OGTT involved two blood glucose tests at fasting state and after 2 hours; in the fasting state, a 75g glucose load is taken either at home or in a collection center. The study showed generally positive results. However, this study is somewhat dated having been conducted more than 20 years ago, and there is a lack of recent research on understanding women's attitudes or experiences of GDM screening, especially using the new IADPSG universal one-step screening approach for GDM, which is globally recommended and increasingly adopted. Under the IADPSG approach, the OGTT involved three blood glucose tests at fasting state, after one hour, and after two hours (Panel, 2010). Evidences showed that the new IADPSG approach has the benefit of overlooking fewer women in probable need of treatment (HAPO, 2008; Lapolla *et al.*, 2011; Benhalima *et al.*, 2013) and enabling diagnosis during the course of one visit rather than two for the GDM women (the minority) as compared to the two-step screening approach (Shang & Ma, 2011; Wei & Yang, 2011). However, under the new approach, a more complicated OGTT test (fasting, two hours, three blood samples required) has to be carried out for all pregnant women. Earlier, the non-GDM pregnant women (the majority) needed only a relatively simple 50g GCT screening test (non-fasting, one hour, one blood sample required) under the two-step screen approach or needed only a simpler OGTT (two blood samples required) under the older OGTT criteria.

There is gap in literature as to what pregnant women think of the IADPSG screening approach. Therefore, it is necessary to explore women's perspectives on the one-step universal approach for GDM in order to fill this gap and to provide evidence for identifying and optimising the best GDM screening approach.

1.4 RATIONAL OF THE RESEARCH

1.4.1 A summary of the current situation

Inconsistent recommendations and research evidences exist as to whether universal screening or selective screening is the best screening approach. The interest of screening for patients with no risk factors remains controversial; however, if only screening the high risk women, low sensitivities and specificities have been reported. Clear consensus does not present on whether or not to adopt the IADPSG one-step screening approach for GDM. There is no RCT evidence for the intervention of the IADPSG criteria, although several observational studies have shown the clinical effectiveness of this approach. However, the new IADPSG approach results in increased costs for the healthcare system and greater burden for non-GDM pregnant women. No study has explored the attitudes or experiences of pregnant women with reference to the new IADPSG approach, thereby failing to provide a clear user perspective.

1.4.2 Gaps and implications for further research and practice

When there is no conclusive answer as to the optimal GDM screening approaches, countries might be suggested to continue with their current practice before future evidence emerges. Compared to universal screening, the selective screening

approach could be beneficial in terms of reduced testing, less cost, more cost-effective and better risk stratification (Hieronimus & Le Meaux, 2010; Waugh *et al*., 2010). Reduced testing would mean that patients would experience potentially less anxiety and discomfort. At the same time, better cost-effectiveness and risk stratification would allow resources to be better allocated to the care needed the most by the patients. However, this strategy also has the disadvantage of overlooking a certain proportion of GDM cases as it routinely excludes low risk women from the screening test by operating under the assumption that they will not have GDM (Scott *et al.*, 2002). Hence, assessment of the effectiveness and cost-effectiveness of selective screening as well as the best selection criteria for identifying high risk women are urgently needed. In this regard, a cutting-edge comparative systematic review might be needed to provide high quality evidence for addressing this issue under discussion.

The new IADPSG one-step universal screening approach for GDM has been recommended since 2010 (Panel, 2010). Recent studies have also favourably appraised the IADPSG criteria in terms of short-term clinical outcomes (Lapolla *et al.*, 2011; Shang & Ma, 2011; Lu *et al*., 2012; Benhalima *et al.*, 2013; Jiang *et al.*, 2013). However, current evidence is still limited, reflecting a lack of evaluation as to the cost-effectiveness of the approach or insights into user perspectives. The adoption of the new approach has yet to be fully investigated. Changes to the current approach might need to be made if new evidence comes into sight.

1.4.3 Rationale of the current PhD research

The current PhD research aimed to assess and explore the optimal screening approaches for GDM in order i) to address the uncertainty of current evidence and

controversies of implementations in GDM screening approaches, and ii) to fill the research gap in terms of the IADPSG screening approach.

This study used a mixed method approach to address the research question. Firstly, a systematic review was undertaken to synthesise evidence on the effectiveness and cost-effectiveness of universal versus selective screening. This aimed to appraise and evaluate the relevant literature and to identify further research needs. Secondly, the perspectives of pregnant women about GDM screening were investigated using Q methodology to address the limited evidence available on user viewpoints. Specifically, the study aimed to provide timely evidence of pregnant women's attitudes towards the new IADPSG approach being adopted. Thirdly, to address the controversy between the clinical effectiveness and the increased cost of the IADPSG approach, the final study explored and assessed a risk score-based selective screening approach for IADPSG.

1.5 Summary

GDM prevalence has been rising with the increasing global epidemic of obesity and type 2 diabetes, as well as the increasing age of conception and child bearing. Screening for GDM after 24th week of gestation is beneficial in reducing maternal and neonatal complications. There is no agreement as to the best screening approaches for GDM. Controversies exist in whether universal screening or selective screening should be recommended. A comprehensive systematic comparative review of selective screening versus universal screening will help to address the issue. Although the adoption of the new IADPSG one-step universal approach for GDM is widely debated, China has been implementing the IADPSG approach since 2011. Therefore, it is worthwhile to investigate the adoption of the IADPSG approach as

well as user attitudes in China where research gaps in the literature related to the screening approach exist.

Chapter 2：Aims and Objectives

2.1 Research question

The main purpose of this thesis is to address the inconsistent evidence and research gap on selective versus universal screening for GDM as well as the adoption of the new IADPSG universal screening approach for GDM. The overall research question framing the research was to evaluate and explore the most appropriate screening approach for GDM.

2.2 Aims

1) To comparatively assess selective screening versus universal screening approach for GDM.
2) To explore patient perspectives on the new IADPSG universal screening approach for GDM.
2) To establish and evaluate a risk score-based selective screening approach under the new IADPSG criteria for GDM.

2.3 Objectives

1) To assess the effectiveness and cost-effectiveness of selective screening in comparison with universal screening for GDM.
2) To explore the pregnant women's attitudes, views, and experiences of the IADPSG universal screening approach for GDM within a Chinese context.
3) To investigate the risk factors for GDM and establish a risk scoring algorithm for the identification of high risk women for a selective screening approach under the new IADPSG criteria within a Chinese context.
4) To evaluate the effectiveness of the risk score-based selective screening approach under the new IADPSG criteria within a Chinese context.

Chapter 3: A Systematic Review of the Effectiveness and Cost-effectiveness of Screening for GDM: Universal or Selective Screening?

3.1 BACKGROUND

GDM has been defined as the onset or first recognition of glucose intolerance during pregnancy (IWCGDM, 1985). Screening and diagnosis for GDM are usually conducted in pregnant women between the 24th to 28th weeks of gestation. There has been debate over the best screening approach for GDM, and it remains unclear whether all pregnant women should be screened through universal screening or only those women at high risk should be screened selectively. The effectiveness and cost-effectiveness of the two approaches are still unclear and require a systematic synthesis of the current evidence.

3.1.1 GDM screening and the controversies

The prevalence of GDM has increased over the last 20 years, and ethnic differences have been reported (Ferrara, 2007). The U.S. Preventive Services Task Force (USPSTF) concluded that there was a moderate net benefit of screening for GDM after the 24th weeks of gestation to reduce maternal and fetal complications (Moyer, 2014). As described in Chapter 1, consensus does not exist as to whether universal or selective screening should be recommended, and different countries continue to implement different screening approaches worldwide. Universal screening requires all pregnant women to undergo the GDM test. On the other hand, selective screening selects pregnant women who are at high risk of developing GDM based on existence of risk factors, and only these high risk women undergo the GDM test. The commonly used GDM risk factors are advanced maternal age, obesity, family history of diabetes (amongst first-degree family members), history of GDM, history of macrosomia, and certain ethnic/racial group including South Asia (especially India, Pakistan or Bangladesh) (Association, 1997; Force, 2008; Walker, 2008; Association, 2010; NICE, 2015).

3.1.2 Current state of evidence on universal versus selective screening

Two systematic reviews (Hieronimus & Le Meaux, 2010; Tieu *et al.,* 2010) and two health technology assessment (HTA) reports (Scott *et al.,* 2002; Waugh *et al.,* 2010) have been identified, which compare the effectiveness of selective screening versus universal screening for GDM.

Hieronimus and Le Meaux (2010) searched only two databases of Medline and Cochrane database from 1990 to 2010 and did not assess the quality of the included studies. Tieu *et al.* (2010) evaluated the screening, diagnosis and treatment of GDM, including only randomised and quasi-randomised trials. Only one quasi-experimental study (Griffin *et al.,* 2000) was identified that compared the two screening approaches, which suggested universal screening was superior to selective screening and that it resulted in higher detection rate, earlier diagnosis and better pregnancy outcomes. However, due to the limitations of this trial, the systematic review by Tieu *et al.* could not arrive at a definite conclusion about which screening approach could be recommended.

Scott *et al.* (2002) and Waugh *et al.* (2010) conducted Health Technology Assessment (HTA) on the treatment and screening for GDM. The review by Scott *et al.* had a particular focus on screening methods and costs as well as an appraisal of GDM screening against the criteria set by the UK National Screening Committee (NSC). Waugh *et al.* (2010) updated the report by Scott *et al.* by examining evidence that has emerged since 2002, including the Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS), the Maternal and Fetal Medicine Units Network (MFMUN) trial and the HAPO and reviewed the trends in maternal age and obesity and their effect on GDM prevalence. It is important to point out that the reports by Scott *et al.* (2002) and Waugh *et al.* (2010) did not focus specifically on

universal screening versus selective screening approaches. Within each report, effectiveness studies on selective screening (or risk factor screening) were subsumed under the section reviewing different GDM screening tests. Hence, these were limited in number as the analysis was peripheral to the main focus of these two studies.

3.1.3 Rationale for conducting the systematic review

The previous two systematic reviews (Hieronimus & Le Meaux, 2010; Tieu *et al.*, 2010) and two HTA reports (Scott *et al.*, 2002; Waugh *et al.*, 2010) were limited for a number of reasons. In the case of Hieronimus & Le Meaux (2010), the limitation was due to the databases that had been searched, whereas the types of study included in Tieu *et al.* (2010) were restricted. The key limitation of Scott *et al.* (2002) and Waugh *et al.* (2010) resided in the focus and quantity of the effectiveness studies included therein.

These four reviews also used different outcome measures. Hieronimus and Le Meaux (2010) used efficacy outcome measures of sensitivity and specificity for the effectiveness studies. In the review by Tieu *et al.*, (2010), clinical outcome measures of key maternal and neonatal events for the effectiveness studies were used. While Scott *et al.* (2002) and Waugh *et al.* (2010) narratively described the sensitivity and specificity for the effectiveness studies involved, as a short-section within the HTA reports.

The current systematic review extends and updates the previous reviews by using broader types of study design, including recent studies published after 2010 and by fully searching and assessing the efficacy outcome measures of sensitivity and specificity of selective screening as compared to universal screening. This systematic

review aimed to produce most up-to-date and comprehensive synthesis of existing evidence on selective versus universal screening approach. Data on sensitivity and specificity were statistically analysed and meta-analysis was conducted where applicable. The findings of the review provide more comprehensive and strong evidence for recommendations as to GDM screening approaches, in order to address the global debate and lack of consensus over selective versus universal screening.

3.2 AIM AND OBJECTIVES

The study aimed to undertake a systematic review on the effectiveness and cost-effectiveness of selective screening compared to universal screening for pregnant women.

The specific objectives were:
(1) To evaluate the effectiveness (sensitivity and specificity) of selective screening compared to universal screening for GDM;
(2) To assess the cost-effectiveness and cost of selective screening compared to universal screening for GDM.

3.3 RESEARCH METHODOLOGY

3.3.1 General framework of the systematic review

The systematic review adhered to the PRISMA guidelines (Moher *et al*., 2009). The protocol of the systematic review was registered with the PROSPERO international prospective register of systematic reviews. The registration number is

CRD42013004241. The proposal can be accessed via: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013004241#.VPOBRPmsWe4.

3.3.2 Search strategy

A wide range of electronic databases including Medline, Embase, Web of Science (SCI and SSCI) and Cochrane Database were searched. Databases of ongoing trials were also searched. These included current Controlled Trials, ClinicalTrials.gov, UKCRN Portfolio Database, and WHO International Clinical Trials Registry Platform.

Advanced searches were deployed with keywords being mapped into subject headings (search terms). Key words were defined from the review question, such as "gestational diabetes" and "screening". It was not necessary to define the study design, outcome or setting at this early stage. Study design and outcome measure were scrutinised at the later screening stage by the researcher, and there was no restriction in terms of the study setting.

The search was restricted to the English language and the publication year (1980 to present). Year 1980 was chosen because the screening approach was recommended at the First Workshop on GDM in 1980 (ADA, 1980). Universal screening was recommended by the first three workshops (Association, 1980; Association, 1985; Metzger, 1991), and selective screening was recommended by the fourth and fifth workshop (Metzger & Coustan, 1998; Metzger *et al.*, 2007). The search was undertaken on 14 November 2014. The search history based on the four databases is shown in Appendix 1.

'Grey literature' was searched to reduce publication bias. These involved unpublished conference proceedings, guidelines, and other information from key relevant organisations in the field of GDM, i.e., ADA (American Diabetes Association), British Diabetes Association (BDA) and the National Institute for Health and Care Excellence (NICE) in the UK. An ancestry search was conducted to check reference lists of relevant articles.

3.3.3 Selection criteria (inclusion and exclusion) for eligible studies

Articles identified in the above search were screened according to the selection criteria (inclusion and exclusion) in the order of title, abstract and full text of studies. In view of the research question framing this systematic review, the selection criteria were developed. These included the types of study, intervention, comparator, outcome, population and setting. Table 3 shows the inclusion and exclusion criteria for selection of studies.

Table 3. A table of inclusion and exclusion criteria for study selection

	Inclusion	**Exclusion**
Study design	1. For effectiveness: RCT, quasi-RCT, cohort study, cross-sectional study 2. For cost-effectiveness and cost: cost-effectiveness study, cost study	Case-control study (not suitable for the research question) Qualitative study Systematic review Literature review
Population	Pregnant women	Women who have pre-existing diabetes or who have already been diagnosed as having GDM before screening
Intervention	Selective screening for GDM (screening can be screening test followed by diagnosis test or one-step diagnosis test)	Screening only contained screening test without diagnosis test

Comparison	Universal screening for GDM (screening can be screening test followed by diagnosis test or one-step diagnosis test)	Screening only contained screening test without diagnosis test
Outcome	1. For effectiveness: (1) sensitivity (percentage of GDM women who were correctly diagnosed as having GDM), (2) specificity (percentage of non-GDM women who avoided GDM screening) 2. For cost and cost-effectiveness: (1) cost studies: cost per identified GDM case; (2) cost-effectiveness: ICER	Studies failed to report the outcome indicated in the inclusion criteria
Setting	Any setting	None

(1) Types of studies: For assessing the effectiveness studies, RCT, quasi-RCT, cohort study, and cross-sectional study were included. Case-control study design was not suitable for the research question, and was excluded. For assessing cost-effectiveness, cost-effectiveness and cost studies were included. Qualitative study, systematic review, and literature review were excluded.

(2) Types of population and the setting: Pregnant women were the target population. However, studies in which women had pre-existing diabetes or had already been diagnosed as having GDM before screening were excluded. No limit was made on population size or study setting.

(3) Types of interventions and comparators: The intervention was the selective screening approach, whereas the comparator was universal screening approach. The GDM test for the two approaches could be either two-step tests (screening test followed by diagnostic test) or one-step test (diagnostic test). Studies which only involved a screening test without a diagnostic test would be excluded.

(4) Types of outcome measures: For effectiveness studies, the outcome measures were sensitivity and specificity of selective screening compared to universal screening. For this review, sensitivity is the proportion of GDM cases who are correctly screened and diagnosed; specificity is the proportion of non-GDM women who can be exempted from the screening and diagnosis. For cost-effectiveness studies, the outcome measure was incremental cost-effectiveness ratio (ICER); whereas for cost studies, the outcome measure was cost per identified GDM case.

3.3.4 Study selection process

All retrieved records were entered into an Endnote database. Duplicate records were identified and removed. Two reviewers pilot-tested a priori screening form based on the predefined study eligibility criteria. The two reviewers screened 20% of all identified records for title and abstract and then all records for full text independently, against the agreed inclusion and exclusion criteria. Level of agreement was tested by the Kappa inter-rater reliability comparison. Disagreements over eligibility were resolved through discussion between the two reviewers to make consensus or by consulting the author of the original studies. The selection process was documented using a PRISMA diagram. Reasons for exclusion of full text papers were documented.

3.3.5 Quality assessment strategy

For effectiveness studies, the Downs and Black quality assessment tool was used (Downs & Black, 1998), which is suitable for both randomised and non-randomised studies. The assessments include quality of reporting, power, internal validity (bias and confounding) and external validity (Downs & Black, 1998). The original Downs and Black checklist was modified as not all items were relevant to the effectiveness

studies in this review. The details of the amended Downs and Black checklist are provided in Appendix 4.1. For cost-effectiveness studies, Drummond checklist was used (Drummond, 1996).

3.3.6 Data extraction strategy

Data extraction forms were developed as informed by the PRISMA guidelines (Moher *et al*., 2009). The relevant data were extracted from included studies by the main reviewer. A second reviewer checked the data extraction result of one included study; uncertainty and/or any disagreements were resolved by discussion. The extracted data of each study were entered into a summary table (Table 4 and Table 5) and a full data extraction table (Appendix 5). The extracted data included the following:

- Study characteristics (i.e., author's name, country and city, study design, study setting, sample size, time of study, funding source)
- Patient baseline characteristics (i.e., inclusion/exclusion criteria, number of enrolled/ analysed participants, body mass index, age, race, family history of diabetes)
- Intervention/ comparator characteristics (i.e., selection criteria for high risk women, time and method of GDM test)
- Outcome characteristics (i.e., sensitivity, specificity; cost per GDM diagnosed; ICER)

For effectiveness studies where sensitivity and specificity were not directly reported but could be calculated from the data, or where sensitivity and specificity were calculated using different definitions, the calculations were done by the reviewer using the definitions in the review.

3.3.7 Data synthesis and interpretation

For effectiveness studies, variables in each study were summarised in text and summary tables to observe the similarities and differences. Meta-analysis of sensitivity and specificity was conducted in Revman 5; forest plots were produced. Pooled estimate of sensitivity and specificity as well as heterogeneity of the included studies were analysed by the HSROC (hierarchical summary receiver operating characteristic) curve in STATA 12.0. Positive predictive values (PPV) and negative predictive values (NPV) according to GDM prevalence were calculated in MS Excel. Publication bias of the included studies was evaluated using the Harbords method for assessing small study bias using STATA 12.0.

3.4 RESULTS

3.4.1 Included studies

A total of 3577 records were identified, with the removal of duplicates reducing this number to 2943 records. Of these, 2837 records were excluded as irrelevant at title and abstract stage, leaving behind 97 potentially relevant records. Of the 97 full text records screened, 64 records were excluded, and the remaining 33 records were included in the review. Amongst the 33 studies selected, there were 28 effectiveness studies, one cost-effectiveness study and four cost studies. The PRISMA flow chart outlining the process of identifying relevant literature is shown in Figure 3. A list of included studies can be found in Appendix 2. The kappa analysis of 20% articles (570 articles) at title and abstract stage achieved a k-value of 0.70, whereas the kappa analysis of 20% articles (6 articles) at full text stage achieved a k-value of 1.00. These outcomes achieved substantial (0.61-0.80) and perfect (0.81-1) agreement (Gwet, 2010).

A list of records excluded at full text stage with reasons for exclusion is provided in Appendix 3. The main reasons for exclusion are broadly outlined herein. Twenty studies had only abstracts or access to full text articles was not available. Nineteen studies were excluded because of outcome measures, as these studies failed to report either sensitivity or specificity according to the definition in the review. Further, the same could not be calculated from their data. Fifteen studies were excluded because they used study designs that fell under the exclusion criteria. Four studies were excluded because of the intervention/comparator, as they did not compare selective screening with universal screening. Three studies were excluded because of intervention/comparator, as they performed two-step tests for the universal screening and one-step test for the selective screening and therefore did not offer comparability. Another two studies were excluded as duplicate publications of an already included study, and one study was excluded because of intervention/comparator, as the study did not include the GDM diagnostic test.

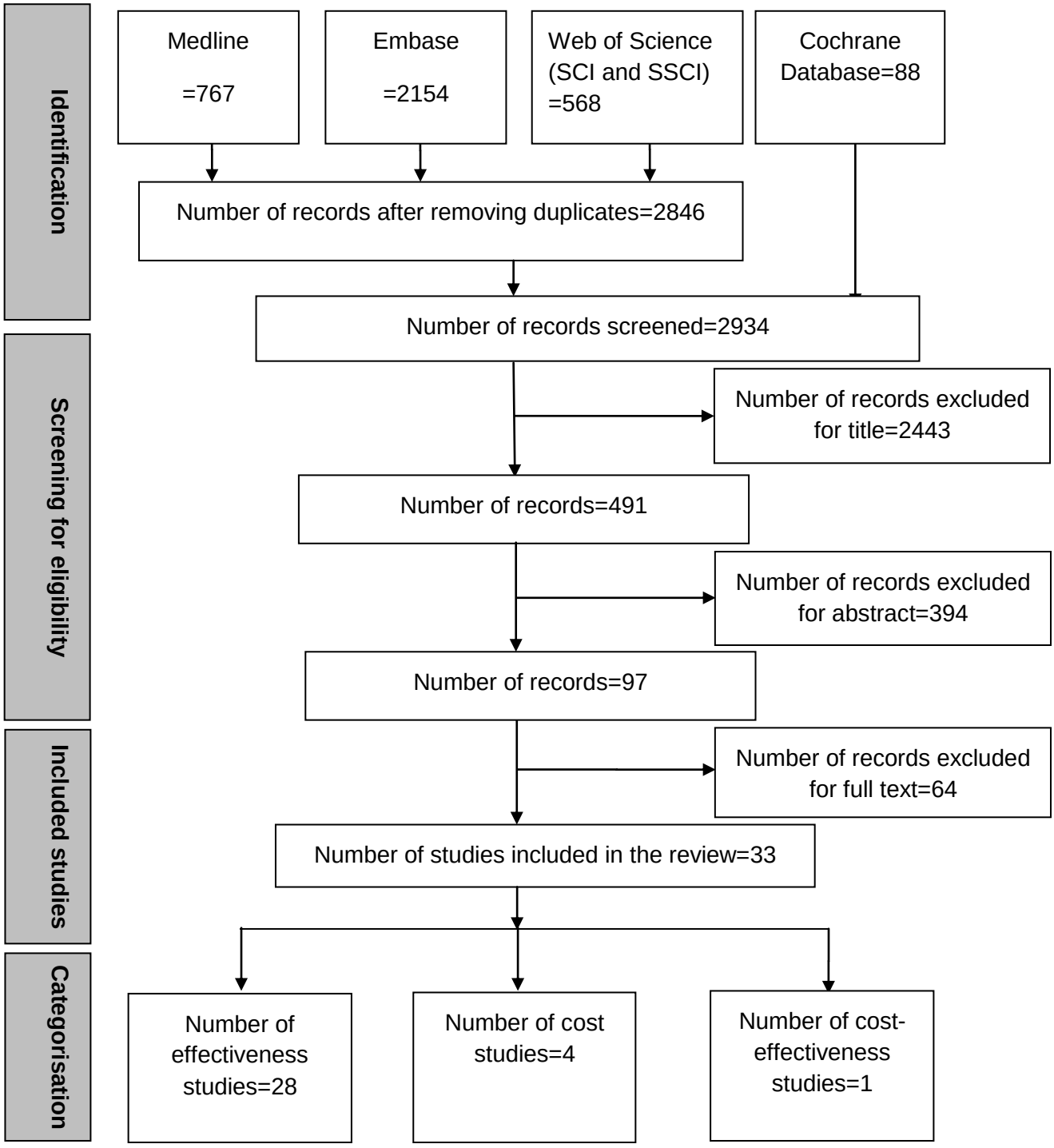


Figure 3. PRISMA flow chart of the study selection process

3.4.2 Result of the effectiveness studies

3.4.2.1 Quality assessment result of the effectiveness studies

The Downs and Black checklist was used for assessing quality of the effectiveness studies. The checklist can be accessed through: http://jech.bmj.com/content/52/6/377.full.pdf+html. The checklist was modified to suit the effectiveness studies in this review. The modified version of the checklist consisted of 16 items with a maximum score of 16 points.

The 28 effectiveness studies achieved an average score of 13.4, ranging from 11 to 15 out of 16 points. The most frequently failed items were due to a number of reasons: (1) 27 studies failed to report the sample calculation (item 27); (2) Eight studies were unable to guarantee that the participants were representative of the entire population because of more than 10% exclusion rate (due to refusal or unavailable data) (item 12); (3) Six studies failed to report the outcomes to be measured clearly in the Introduction or Methods section (item 2); and (4) Six studies failed to take into account the loss to follow-up (if exceeding 10%) in their analysis (item 26). The remaining items were generally well fulfilled by the studies. The assessment result is illustrated in Appendix 4.2.

3.4.2.2 Characteristics of the effectiveness studies

Data extraction for study characteristics of the 28 effectiveness studies is available in Appendix 1. A summary containing the key characteristics of the 15 studies with two-step GDM tests (50g GCT followed by OGTT) and 13 studies with a one-step GDM test (75g/100g OGTT) is shown in Table 4 and Table 5.

(1) 15 studies with two-step GDM tests

The sample size of these studies ranged from 363 (Zoller *et al.,* 1988) to 25118 (Williams *et al.,* 1999). Seven studies were conducted in the USA (Coustan *et al.,* 1989; Danilenko-Dixon *et al.,* 1999; Helton *et al.,* 1997; Lavin *et al.,* 1985; Sacks *et al.,* 1987; Williams *et al.*, 1999; Zoller *et al.*, 1988). Four studies were conducted in European countries, from Spain (Jimenez-Moleon *et al.,* 2002), Italy (Di Cianni *et al.,* 2003), Netherlands (Van Leeuwen *et al.,* 2010), and Turkey (Caliskan *et al.,* 2014), respectively. The remaining four studies were conducted in Australia (Teh *et al.,* 2011), Canada (Naylor *et al.,* 1997), Thailand (Arora, *et al.,* 2013), and Iran (Hadaegh *et al.,* 2005), respectively.

Excepting one cross-sectional study (Arora *et al.,* 2013), the other 14 studies were cohort studies. This involved six prospective studies (Coustan *et al.,* 1989; Hadaegh *et al.,* 2005; Lavin *et al.,* 1985; Sacks *et al.,* 1987; Van Leeuwen *et al.,* 2010; Zoller *et al.,* 1988), the other eight studies were retrospective studies.

The selection criteria for high risk women they used varied from as few as three risk factors of age, obesity, and family history of diabetes (Di Cianni *et al.,* 2003; Moses *et al.,* 1995) or age, obesity, and ethnicity (Moses *et al.,* 1998; Naylor *et al.,* 1997) to as many as ten risk factors (Jimenez-Moleon *et al.,* 2002). The details of the selection criteria risk factors are also available in Table 4.

(2) 13 studies with a one-step GDM test

The sample size of these studies ranged from 768 (Shamsuddin *et al.,* 2001) to 18775 (Cosson *et al.,* 2013). Seven studies were conducted in European countries, i.e., three in Italy (Capula *et al.,* 2013; Corrado *et al.,* 2014; Pintaudi *et al.,* 2014), one in France (Cosson *et al*., 2013), one in Sweden (Ostlund & Hanson, 2003), one in Denmark (Jensen *et al.,* 2003), and one in eleven Mediterranean countries (Savona-Ventura *et al.,* 2013). Two studies were conducted in Australia (Moses *et al.,* 1995; Moses *et al.,* 1998). The remaining four studies were conducted in Iran (Shirazian *et al.,* 2009), Singapore (Chong *et al.,* 2014), Malaysia (Shamsuddin *et al.,* 2001), and Sri Lanka (Wagaarachchi *et al.,* 2001), respectively.

The 13 studies involved six prospective cohort studies (Chong *et al.,* 2014; Jensen *et al.,* 2003; Moses *et al.,* 1995; Ostlund & Hanson, 2013; Savona-Ventura *et al.,* 2013; Wagaarachchi *et al.,* 2001), six retrospective cohort studies (Capula *et al.,* 2013; Corrado *et al.,* 2014; Cosson *et al*., 2013; Moses *et al.,* 1998; Pintaudi *et al.,* 2014; Shirazian *et al.,* 2009), and one cross-sectional studies (Shamsuddin *et al.,* 2001). The sample size ranged from 768 women (Shamsuddin *et al.,* 2001) to 18775 women (Cosson *et al.,* 2013).

The selection criteria for high risk women they used varied, from as few as three risk factors of age, obesity, and family history of diabetes (Moses *et al*., 1995) or age, obesity, and ethnicity (Moses *et al.,* 1998), to as many as 9 risk factors (Shamsuddin *et al*., 2001). The details of the selection criteria risk factors are also available in Table 5.

Table 4. Summary table of key characteristics of the 15 effectiveness studies with two-step GDM tests

Author/ year/ Country	Type of study/ Number of participants	Screening criteria/ Diagnosis criteria	GDM prevalence by universal/ selective screening	Selective screening criteria for high risk women								Specificity	Sensitivity	Author's conclusion
				Age	Obesity (BMI)	Family history of diabetes	Personal history of GDM	A prior macrosomic infant	A history of adverse obstetric outcome	Ethnicity	Others			
Arora/ 2013 Thailand	Cross-sectional study/ 593	50g GCT/ 100g OGTT	9.3%/ 7.3%	≥30	≥25	√	√	√	√	—	glucosuria, hypertension	47.2%	78.2%	Selective screening might not be acceptable and needs re-evaluation
Caliskan/ 2014 Turkey	Retrospective cohort study/ 422	50g GCT/ 100g OGTT	3.3%/ 3.3%	≥25	≥25	√	—	√	√	—	—	30.0%	100.0%	Recommend selective screening
Coustan/ 1989 USA	Prospective cohort study/ 6214	50g GCT/ 100g OGTT	2.0%/ 1.5%	≥30	≥85th percentile for height	√	√	√	√	—	—	56.0%	65.0%	Do not recommend selective screening
Danilenko-Dixon/ 1999 USA	Retrospective cohort study/ 18504	50g GCT/ 100g OGTT	3.0%/ 3.0%	≥25	≥27	√	—	—	—	√	—	9.9%	97.0%	Do not recommend selective screening
Di Cianni/ 2003 Italy	Retrospective cohort study/ 3950	50g GCT/ 100g OGTT	8.1%/ 8.0%	≥25	≥25	√	—	—	—	—	—	5.6%	98.4%	Recommend universal screening

Hadaegh/ 2005 Iran	Prospective cohort study/ 800	50g GCT/ 100g OGTT	8.9%/ 7.9%	≥25	≥25	√	—	—	—	—	—	31.0%	88.7%	Recommend universal screening
Helton/ 1997 USA	Retrospective cohort study/ 3950	50g GCT/ 100g OGTT	2.5%/ 1.7%	≥35	≥200lb	√	√	√	√	—	—	74.6%	69.2%	Recommend selective screening
Jimenez-Moleon/ 2002 Spain	Retrospective cohort study/ 2574	50g GCT/ 100g OGTT	2.5%/ 2.3%	Age≥30 (ACOG) or ≥25 (ADA)	≥27	√	√	√	√	—	chronic hypertension, polyhydramnios, hpertension induced by the pregnancy, suspected large fetus for gestational age	44.2% (ACOG critieria); 15.5% (ADA criteria)	89.2% (ACOG criteria); 96.9% (ADA criteria)	Selective screening is desirable only when fairly restrictive criteria are applied in defining the gravidae at risk
Lavin/ 1985 USA	Prospective cohort study/ 2077	50g GCT/ 100g OGTT	1.4%/ 0.7%		√	√	√	√	√	—	monilial vaginitis, glucosuria, polyhydramnios, an infant suspected of being large for gestational age	53.8%	46.7%	Do not recommend selective screening
Naylor/ 1997 Canada	Retrospective cohort study/ 3131	50g GCT/ 100g OGTT	2.1%/ 1.9%	≥30	≥22	√	—	—	—	—	—	34.7%	90.6%	Recommend selective screening
Sacks/ 1987 USA	Prospective cohort study/ 4116	50g GCT/ 75g OGTT	3.4%/3.3%	≥25	≥150pounds	√	—	√	√	—	—	23.0%	97.1%	Recommend selective screening in early pregnancy
Teh/ 2011 Australia	Retrospective cohort study/ 2880	75g GCT/ 75g OGTT	8.7%/8.1% (NICE criteria); 8.7% (ADA criteria); 8.6% (ADIPS criteria)	NICE criteria								30.1% (NICE criteria); 3.5% (ADA criteria); 12.6% (ADIPS criteria)	92.7% (NICE criteria); 100% (ADA criteria); 98.6% (ADIPS criteria)	Do not recommend selective screening
					≥30	√	√	√	—	√	—			
				ADA criteria										
				≥25	abnormal weight	√	—	—	√	√	history of abnormal glucose metabolism			
				ADIPS criteria										
				≥30	N/A	√	√	—	√	√	glycosuria			

Van Leeuwen/ 2010 Netherlands	Prospective cohort study/ 995	Random plasma glucose and 50g GCT/ 75g OGTT	4.6%/ 3.5%		linear relationship between 22-30	√	√	—	—	√	—	57.0%	75.0%	Recommend selective screening
Williams/ 1999 USA	Retrospective cohort study/ 25118	50g GCT/ 100g OGTT	0.8%/ 0.8%	≥25	≥27	√	—	—	—	√	—	11.1%	96.0%	No direct conclusion (the figures didn't support selective screening from the reviewer's view)
Zoller/ 1988 USA	Prospective cohort study/ 363	50g GCT/ 100g OGTT	2.8%/ 1.1%		≥90.72 kg	√	√	√	√	—	glucosuria, polyhydramnios, intrauterine growth consistent with large gestational aged infant	61.4%	40.0%	Recommend universal screening

Table 5. Summary table of key characteristics of the 13 effectiveness studies with a one-step GDM test

Author/ year/ Country	Type of study/ Number of participants	Screening criteria/ Diagnosis criteria	GDM prevalence by universal/ selective screening	Selective screening criteria for high risk women								Specificity	Sensitivity	Author's conclusion
				Age	Obesity	Family history of diabetes	Personal history of GDM	A prior macrosomic infant	A history of adverse obstetric outcome	Ethnicity	Others			
Capula/ 2013 Italy	Retrospective cohort study/ 2448	75g OGTT (IADPSG)	27.5%/ 20.5%	≥35	≥25	√	√	√	—	√	—	22.0%	74.6%	Recommend universal screening
Chong/ 2014 Singapore	Prospective cohort study/ 1136	75g OGTT (WHO 1999)	18.9%/ 9.8%		>30	√	√	√	—	√	—	56.3%	51.6%	Recommend universal screening
Corrado/ 2014 Italy	Retrospective cohort study/ 1015	75g OGTT (IADPSG)	11.3%/ 8.6%	≥35	≥25	√	√	√	—	—	—	41.7%	77.0%	Need future study
Cosson/ 2013 France	Retrospective cohort study/ 18775	75g OGTT	14.4%/ 9.4%	≥35	≥25	√	√	√	—	—	—	41.5%	65.3%	Do not recommend selective screening
Jensen/ 2003 Denmark	Prospective cohort study/ 5235	75g OGTT	2.4%/ 1.9%		≥27	√	√	√	—	—	Glucosuria	63.7%	80.6%	Recommend selective screening
Moses/ 1995 Australia	Prospective cohort study/ 1185	75g OGTT	6.7%/ 4.1%	≥30	≥30	√	—	—	—	—	—	54.2%	60.8%	Recommend universal screening
Moses/ 1998 Australia	Retrospective cohort study/ 2907	75g OGTT	6.3%/ 5.7%	≥25	≥25	√	—	—	—	—	—	19.7%	91.3%	Selective screening needs further evaluation in different populations

Ostlund & Hanson/2003 Sweden	Prospective cohort study/ 3616	75g OGTT	1.7%/ 1.1%		≥90kg	√	√	√	—	—	—	84.2%	47.5%	Do not recommend selective screening
Pintaudi/ 2014 Italy	Retrospective cohort study/ 1015	75g OGTT (IADPSG)	11.3%/ 10.0%		≥25	√	—	—	—	—	FPG> 4.4mmol/l	36.8%	89.0%	Recommend selective screening using the RECPAM model
Savona-Ventura/ 2013 Eleven Mediterranean countries	Prospective cohort study/ 1368	75g OGTT	8.7%/ 7.0%	≥25	pre-pregnancy BMI ≥25 or a 3rd trimester BMI ≥30	—	—	—	—	—	FBG >5.0 mmol/L	80.7%	65.9%	Recommend selective screening for the areas of economic restraint
Shamsuddin/ 2001 Malaysia	Cross-sectional study/ 768	75g OGTT	24.9%/ 10.1%	≥35	≥80kg	√	√	√	√	—	urinary tract infection, vaginal discharge and pruritis vulvae, glycosuria	33.2%	72.2%	Recommend universal screening
Shirazian/ 2009 Iran	Retrospective cohort study/ 924	75g OGTT	7.4%/ 7.3%	≥25	≥25	√	—	—	—	—	—	13.6%	98.5%	Selective screening do not miss substantial number of GDM cases
Wagaarachchi/ 2001 Sri Lanka	Prospective cohort study/ 1004	75g OGTT	4.1%/ 2.4%	≥35	≥30	√	—	√	√	—	presence of grand multiparity	54.1%	58.5%	Recommend universal screening

3.4.2.3 Data synthesis of the effectiveness studies

The outcomes of GDM incidence, sensitivity, and specificity were summarised for the 15 studies with two-step tests and the 13 studies with a one-step test, respectively.

(1) 15 studies with two-step GDM tests

In this category, the GDM incidence under the universal screening approach ranged from 0.8% in the US (Williams *et al.,* 1999) to 9.3% in Thailand (Arora *et al.,* 2013). Of the 15 studies, five studies recommended selective screening (Caliskan *et al.,* 2014; Helton *et al.,* 1997; Naylor *et al.,* 1997; Sacks *et al.,* 1987; Van Leeuwen *et al.,* 2010; Pintaudi *et al.,* 2014), and 8 studies recommended universal screening. Of the remaining two studies, one recommended selective screening when fairly restrictive selection criteria for risk women were applied (Jimenez-Moleon *et al.,* 2002), and one did not make a direct conclusion (Williams *et al.,* 1999). Figures of sensitivity and specificity of selective screening compared to universal screening for the 15 studies are provided in Table 6. Meta-analysis was conducted in Revman 5, and forest plot of the sensitivity and specificity was produced (Figure 4). The forest plot showed considerable heterogeneity (values) amongst the studies and the trade-off between sensitivity and specificity. The sensitivity ranged from 40.0% (with a specificity of 61.4%) (Zoller *et al.,* 1988), to 100% (with a specificity of 30%) (Caliskan *et al.,* 2014). The specificity ranged from 7.0% (with a sensitivity of 98.7%) (Corcoy *et al.,* 2004), to 74.6% (with a sensitivity of 69.2%) (Helton *et al.,* 1997).

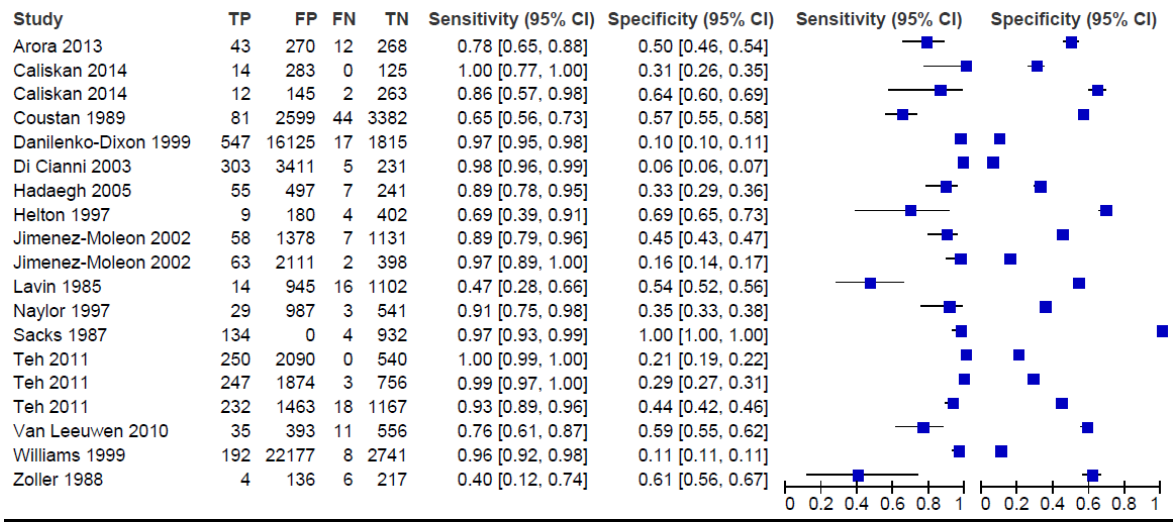


Figure 4. Forest plot of sensitivity and specificity for the 15 studies with two-step GDM tests

There are several reasons for the heterogeneity of sensitivity and specificity between studies, including differences in populations and ethnic groups, different selection criteria for high risk women as well as GDM prevalence. Different populations have different risk profiles for GDM. Table 6 shows that there were three studies which developed own risk criteria (Caliskan *et al.*, 2014) or risk score (Naylor *et al.*, 1997; Van Leeuwen *et al.*, 2010) based on local population characteristics for selecting high risk women for selective screening. All of them found selective screening to be effective and recommended the selective approach to screening. The GDM incidences of the studies which recommended universal screening (ranged from 1.4% to 9.3%; mean 5.5%; median 5.6%) were generally higher than the studies where selective screening (ranged from 2.1% to 4.6%, mean 3.2%; median 3.3%) is recommended.

Table 6. Sensitivity and specificity of selective screening for the 15 studies with two-step GDM tests

Study	Country	GDM incidence (under universal screening)	Use of self-developed selection criteria	Sensitivity	Specificity	Study quality (max score= 16)
Five studies recommending selective screening						
Caliskan/ 2014	Turkey	3.3%	√ (own risk criteria)	100.0%	30.0%	16
Helton/ 1997	USA	2.5%		69.2%	74.6%	12
Naylor/ 1997	Canada	2.1%	√ (own risk score)	90.6%	34.7%	15
Sacks/ 1987	USA	3.4%		97.1%	23.0%	12
Van Leeuwen/ 2010	Netherlands	4.6%	√ (own risk score)	75.0%	57.0%	11
Eight studies recommending universal screening						
Arora/ 2013	Thailand	9.3%	No	78.2%	47.2%	15
Coustan/ 1989	USA	2.0%	No	65.0%	56.0%	13
Danilenko-Dixon/ 1999	USA	3.0%	No	97.0%	9.9%	15
Di Cianni/ 2003	Italy	8.1%	No	98.4%	5.6%	14
Hadaegh/ 2005	Iran	8.9%	No	88.7%	31.0%	15
Lavin/ 1985	USA	1.4%	No	46.7%	53.8%	13
Teh/ 2011	Australia	8.7%	No	92.7% (NICE criteria); 100% (ADA criteria); 98.6% (ADIPS criteria)	30.1% (NICE criteria); 3.5% (ADA criteria); 12.6% (ADIPS criteria)	15
Zoller/ 1988	USA	2.8%	No	40.0%	61.4%	12
Two inconclusive studies						
Jimenez-Moleon/ 2002	Spain	2.5%	No	89.2% (ACOG criteria); 96.9% (ADA criteria)	44.2% (ACOG critieria); 15.5% (ADA criteria)	13
Williams/ 1999	USA	0.8%	No	96.0%	11.1%	12

(2) 13 studies with a one-step GDM test

In this category, the GDM incidence under the universal screening approach ranged from 1.7% in Sweden (Ostlund & Hanson, 2003) to 27.5% in Italy (Capula *et al*., 2013). Of the 13 studies, two studies recommended selective screening (Jensen *et al*., 2013; Pintaudi *et al*., 2014). Seven studies recommended universal screening (Capula *et al*., 2013; Chong *et al*., 2014; Cosson *et al*., 2013; Moses *et al*., 1995; Ostlund & Hanson *et al*., 2003; Shamsuddin *et al*., 2001; Wagaarachchi *et al*., 2001). Of the remaining four studies, one concluded further study was needed (Corrado *et al*., 2014); one recommended selective screening in economic restraint areas (Savona-Ventura *et al.*, 2013); one called for further evaluation (Moses *et al.*, 1988); and one did not make a direct recommendation (Shirazian *et al.*, 2009). Figures of sensitivity and specificity of selective screening compared to universal screening for the 13 studies are provided in Table 7. Meta-analysis was conducted in Revman 5, forest plot of the sensitivity and specificity was produced and provided in Figure 5. Again, the forest plot showed considerable heterogeneity amongst the studies and the trade-off between sensitivity and specificity. The sensitivity ranged from 51.6% (with a specificity of 56.3% (Chong *et al*., 2014), to 98.5% (with a specificity of 13.6%) (Shirazian *et al.* ,2009). The specificity ranged from 13.6% (with a sensitivity of 98.5%) (Shirazian *et al.*, 2009), to 91.3% (with a sensitivity of 19.7%) (Moses *et al*., 1998).

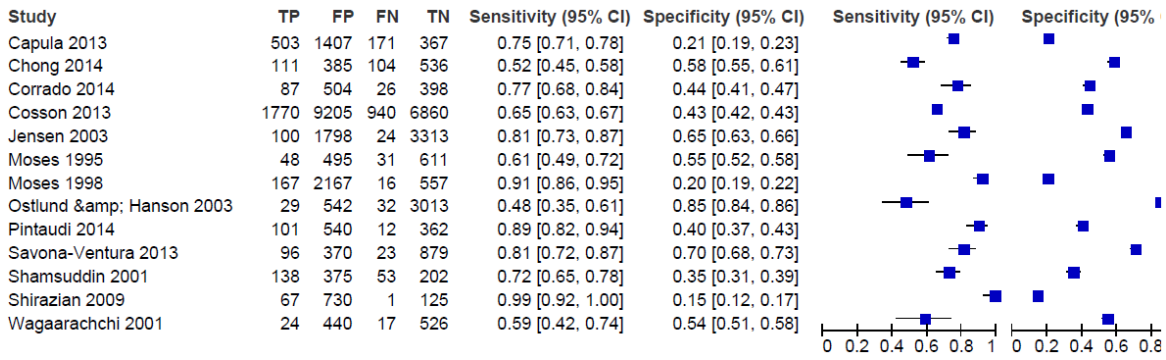


Figure 5. Forest plot of sensitivity and specificity for the 13 studies with a one-step GDM test

Similar to the 15 studies with two-step GDM tests, associations were observed between the effectiveness of the selective screening and the selection criteria and GDM prevalence. As shown in Table 7, the only one study which developed own risk criteria (Pintaudi *et al.*, 2014) for selecting high risk women for selective screening found it was very effective and recommended selective screening. The GDM incidences of the studies which recommended universal screening (ranged from 1.7% to 27.5%; mean 13.6%; median 11.3%) were generally higher than the studies which recommended selective screening (ranged from 2.4% to 11.3%; mean 6.9%; median 6.9%).

Table 7. Sensitivity and specificity of selective screening for the 13 studies with a one-step GDM test

Study	Country	GDM incidence (under universal screening)	Use of own selection criteria	Sensiti-vity	Speci-ficity	Study quality
Two studies recommending selective screening						
Jensen/ 2003	Denmark	2.4%	No	63.7%	80.6%	12
Pintaudi/ 2014	Italy	11.3%	√	89.0%	36.8%	15
Seven studies recommending universal screening						
Capula/ 2013	Italy	27.5%	No	74.6%	22.0%	15
Chong/ 2014	Singapore	18.9%	No	51.6%	56.3%	15
Cosson/ 2013	Italy	11.3%	No	41.5%	65.3%	11
Moses/ 1995	Australia	6.7%	No	54.2%	60.8%	11
Ostlund & Hanson/2003	Sweden	1.7%	No	47.5%	84.2%	13

Shamsuddin/ 2001	Malaysia	24.9%	No	72.2%	33.2%	14
Wagaarachchi/ 2001	Sri Lanka	4.1%	No	58.5%	54.1%	14
Four inconclusive studies						
Corrado/ 2014	Italy	11.3%	No	77.0%	41.7%	14
Moses/ 1998	Australia	6.3%	No	19.7%	91.3%	11
Savona-Ventura/ 2013	Eleven Mediterranean countries	8.7%	No	65.9%	80.7%	13
Shirazian/ 2009	Iran	7.4%	No	98.5%	13.6%	14

(3) Integration of all the 28 effectiveness studies

Combining the two categories, only seven of the 28 effectiveness studies recommended selective screening (Caliskan *et al.,* 2014; Helton *et al.,* 1997; Jensen *et al.,* 2013; Naylor *et al.,* 1997; Sacks *et al.,* 1987; Van Leeuwen *et al.,* 2010; Pintaudi *et al.,* 2014). Fifteen studies recommended universal screening, whereas two studies recommended selective screening conditionally, as in when fairly restrictive criteria are applied in defining the gravidae at risk (Jimenez-Moleon *et al.,* 2002) or for the areas of economic restraint (Savona-Ventura *et al.,* 2013). Two studies concluded that selective screening needed to be assessed in different populations (Moses *et al.,* 1998) or needed further study (Corrado *et al.,* 2014). Two studies did not make a recommendation on universal or selective screening (Shirazian *et al.,* 2009; William *et al.,* 1999).

(4) The heterogeneity of the included studies

The HSROC (hierarchical summary receiver operating characteristic) curve was used to assess the heterogeneity of the studies and pooled estimate of sensitivity and specificity. The HSROC curve was generated using the metandi analysis in STATA 12.0. The results for the 15 studies with two-step tests and 13 studies with a one-step test are shown in Figures 6 and 7, respectively. For both groups, high heterogeneity

was noted in the results of the studies reviewed. For the 15 studies, the pooled estimate of sensitivity and specificity was 0.91 and 0.46, respectively. For the 13 studies, the pooled estimate of sensitivity and specificity was 0.77 and 0.45, respectively.

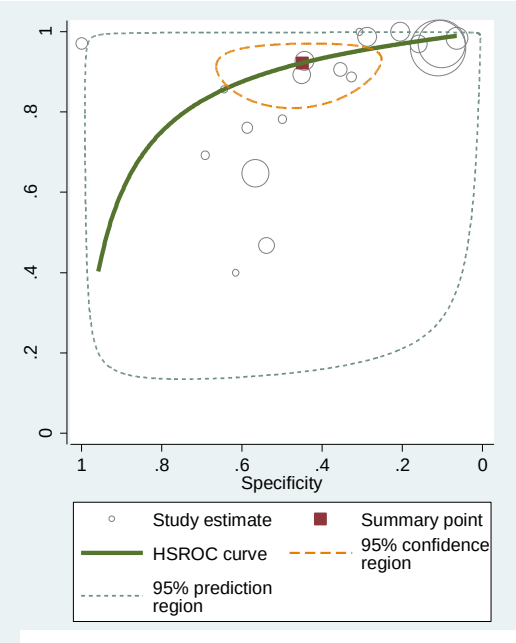


Figure 6. HSROC curve for the 15 studies with two-step GDM test

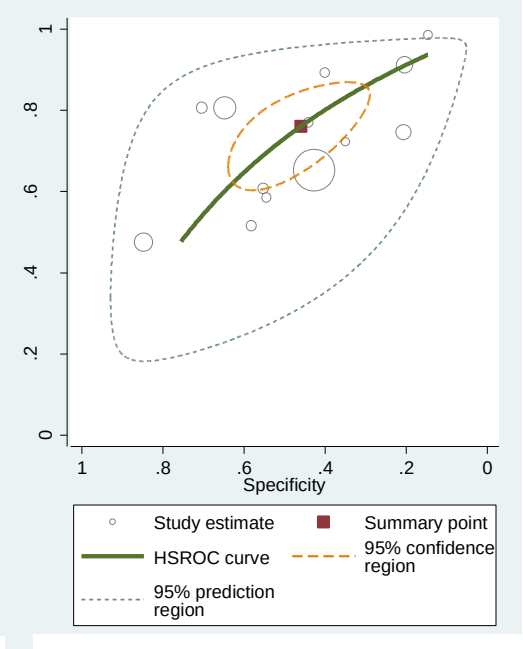


Figure 7. HSROC curve for the 13 studies with a one-step GDM test

(4) The PPV and NPV of the included studies

Although sensitivity and specificity are used as performance indicators for screening or diagnostic test, the positive predictive value (PPV) and negative predictive value (NPV) are often reported. The focus of PPV and NPV is on the predictive power of the screening or diagnostic test. PPV is the probability of disease given that the test is positive. NPV is the probability of no disease given that the test is negative. Figures of PPV and NPV are meaningful to and preferred by clinicians who are interested in using the test result to predict a patient's probability of having (or not having) a disease. However, it should be noticed that PPV and NPV are not intrinsic

to the test, they depend also on the prevalence of the disease (Altman & Bland, 1994). Therefore, they are not suitable for the purpose of serving as performance indicators for a test.

The PPV and NPV were calculated using the pooled sensitivity, pooled specificity and pooled prevalence. The pooled sensitivity and specificity values were generated using STATA 12.0, as shown in the HSROC curve section. The pooled prevalence was generated using MetaAnalyst software, based on the sample size and prevalence of each study. The pooled prevalence of the 15 studies with two-step tests and 13 studies with a one-step test were 3.9% [2.7%, 5.7%] and 10.0% [6.3%, 15.4%], respectively. The PPV and NPV analysis results for the 15 studies with two-step tests and 13 studies with a one-step test are shown in Figures 8 and 9, respectively.

Based on the pooled summary ROC model estimates of sensitivity and specificity, the figures showed the PPV and NPV values according to the prevalence of GDM. As prevalence increases, the PPV increases and NPV decreases. Data points are PPV (red point) and NPV (blue point) at sROC point estimates for sensitivity, specificity and at pooled prevalence. It is worth noticing that the index test here is not the screening test itself (OGTT) but screening women to dichotomise into high or low risk of GDM. For the 15 studies, the PPV and NPV was 6.3% and 99.2%, respectively. This means that a woman has a 6.3% chance of developing GDM if categorised as high risk woman. On the other hand, she has a 99.2% chance of not developing GDM if categorised as low risk woman. For the 13 studies, the PPV and NPV was 13.3% and 94.5%, respectively. This means that a woman has a 13.3% chance of developing GDM if categorised as high risk woman, and a 94.5% chance of not developing GDM if categorised as low risk woman. The ellipses provide a guide to the likely range in PPV and NPV given the meta-analytic pooled estimates for sensitivity and specificity and prevalence; they take into account the 95%

confidence intervals (CIs) of PPV and of NPV at the pooled prevalence and the 95% CI of the pooled predictive values derived from the 95% Lower CI (LCI) and Upper CI (UCL) for the pooled sensitivity and specificity. For the 15 studies and the 13 studies reviewed, the LCI and UCL for PPV were 13.7% and 36.1%, whereas the LCI and UCL for NPV were 91.8% and 95.9%.

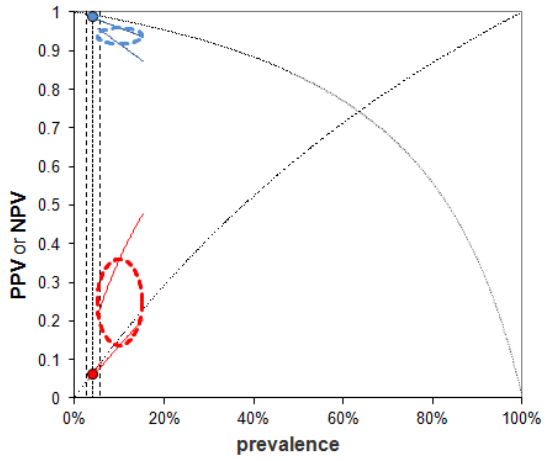


Figure 8. PPV and NPV analysis result for the 15 studies with two-step GDM tests

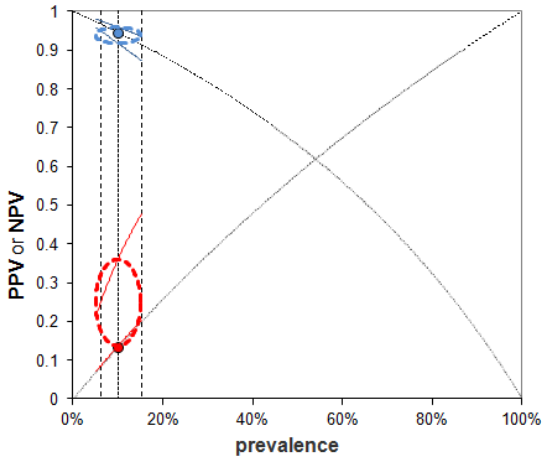


Figure 9. PPV and NPV analysis result for the 13 studies with a one-step GDM

(5) Publication bias analysis of all the 28 effectiveness studies

Harbords method for small study bias was used for assessing publication bias (Category A and B studies together). The results were shown in Figure 10. The estimated intercept is 4.520 with a standard error of 1.694, giving a p-value of 0.012. The modified test suggests there were small-study effects.

```
Note: data input format tcases tnoncases ccases cnoncases assumed.
Note: odds ratios assumed as effect estimate of interest

Harbord's modified test for small-study effects:
Regress Z/sqrt(V) on sqrt(V) where Z is efficient score and V is score variance

Number of studies =  32                                  Root MSE      =   5.899
```

Z/sqrt(V)	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqrt(V)	.1684206	.2721219	0.62	0.541	-.3873264	.7241676
bias	4.504462	1.693561	2.66	0.012	1.045748	7.963175

```
Test of H0: no small-study effects          P = 0.012
```

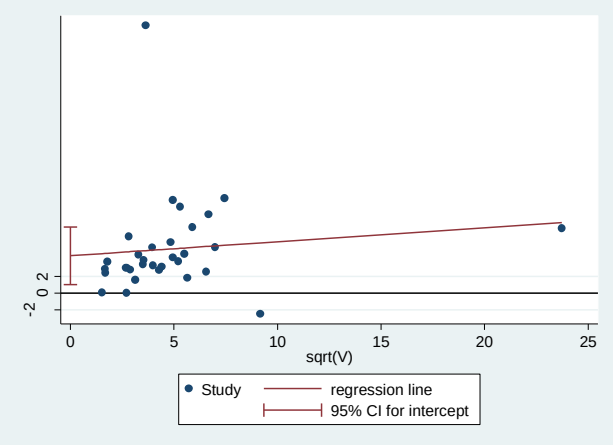


Figure 10. Publication bias analysis for all included studies

3.4.3 Result of the cost-effectiveness and cost studies

3.4.3.1 Quality assessment result of the cost-effectiveness and cost studies

The Drummond checklist was used to assess the quality of cost-effectiveness studies. (Drummond,1996). Poncet *et al.* (2002) fulfilled 23 of the 35 assessment criteria, failed (or not clear) 8, and the other 4 was not appropriate for the study. The assessment result is illustrated in Appendix 4.3.

3.4.3.2 Characteristics of the cost-effectiveness and cost studies

Data extraction result for study characteristics of the cost-effectiveness and cost studies are provided in Table 8 and Table 9. The cost-effectiveness study (Poncet *et al*., 2002) was conducted in France, and used a health system perspective. Of the four cost studies, two were conducted in the USA (Coustan *et al*., 1989; Reed *et al*., 1984), one in Iran (Larijani *et al*., 2004), and one in Malaysia (Shamsuddin *et al*., 2001). All of the included studies used hospital perspective other than the health system perspective. The hospital perspective focuses on the immediate direct costs of screening and diagnosing patients.

Table 8. Summary table of key characteristics of the cost-effectiveness study

Author/ year/ Country	Analytical framework/ Perspective	Intervention/ Comparator	Selective screening criteria for high risk women	Clinical effectiveness measures/ Resource	Cost measures/ Resource	Cost-effectiveness outcomes	Author's conclusion
Poncet/ 2002 France	Decision analysis and cost-effectiveness analysis/ Health system	Screening 1 (S1): High risk women +50g GCT +100g OGTT; Screening 2 (S2): All pregnant women +50g GCT +100g OGTT; Screening 3 (S3): All pregnant women +75g OGTT/	Age≥35, obesity (BMI ≥27), family history of diabetes, personal history of GDM, a prior macrosomic infant, history of adverse obstetric outcome	Macrosomia, prematurity, perinatal mortality, and hypertensive disorders rates/ 38 published articles	Costs of screening tests, obstetrical cares, management of GDM, if any, delivery cares, and sick leave/ A prospective study of 120 pregnant women	ICERs of S2 were 1.10-1.11 times of S1. ICERs of S3 were 3.27-3.75 times of S1.	Favoured selective screening

Table 9. Summary table of key characteristics of the cost studies

Author/ year/ Country and city	Perspective/ Price year	Intervention/ Comparator	Selective screening criteria for high risk women	Clinical effectiveness measures/ Resource	effectiveness outcomes	Cost measures/ Resource	Cost reference	Cost outcomes	Currency	Direct or indirect costs	Discount	Author's conclusion
Coustan/ 1989 USA, Rhode Island	Hospital perspective/ N/A	(1) Universal screening using 50g GCT (140 mg/dl) +100g OGTT; (2) Selective screening A (Age≥25 *et al.*) using 50g GCT (140 mg/dl) +100g OGTT; (3) Selective screening B (Age≥30 *et al.*) using 50g GCT (140 mg/dl) +100g OGTT	Age≥30, obesity (weight ≥85th percentile for height), family history of diabetes, personal history of GDM, a prior macrosomic infant, history of adverse obstetric outcome	Sensitivity/A prospective cohort study of 6214 pregnant women	(1) Universal screening: 100% sensitivity; (2) Selective screening A (Age≥25 *et al.*): 85% sensitivity; (3) Selective screening B (Age≥30 *et al.*): 65% sensitivity	Cost per GDM case detected/A prospective cohort study of 6214 pregant women	N/A	Cost per GDM case detected: (1) $222; (2) $192; (3) $190	US dollars	Direct costs (immediate direct costs incurred by the hospital laboratory in the screenig and diagnosis of GDM)	Discount was not relevant and not reported	Do not recommend selective screening

Author/ year/ Country and city	Perspective/ Price year	Intervention/ Comparator	Selective screening criteria for high risk women	Clinical effectiveness measures/ Resource	effectiveness outcomes	Cost measures/ Resource	Cost reference	Cost outcomes	Currency	Direct or indirect costs	Discount
Larijani/ 2004 Iran, Tehran	Hospital perspective/ 2002	(1) U130 (reference): Universal screening using 50g GCT (130 mg/dl) +100g OGTT; (2) U140: Universal screening using 50g GCT (140 mg/dl) +100g OGTT; (3) S130: Selective screening using 50g GCT (130 mg/dl) +100g OGTT; (4) S140: Selective screening using 50g GCT (140 mg/dl) +100g OGTT	Age≥35, obesity, family history of diabetes, a prior macrosomic infant, history of adverse obstetric outcome, polyhydramnios, glycosuria	GDM prevalences, sensitivities of screening strategies/ A single study of 2416 pregnant women at four university hospitals	GDM prevalences: 4.7% (U130), 4.1% (U140), 4.1% (S130), 3.6 (S140) Sensitivities: 100% (U130), 88% (U140), 86% (S130), 77% (S140)	Cost per pregnant women, cost per GDM case detected/ Same single study of 2416 pregnant women at four university hospitals	Mean value of the public and private sector tariffs, assuming standard material and services	Cost per pregnant women: IRR 30,140 ($3.80) for U130, IRR 25,641 ($3.20) for U140, IRR 21,703 ($2.71) for S130, IRR 19,124 ($2.39) for S140; Cost per GDM case detected: IRR 644,488 ($80.56) for U130, IRR 619,500 ($77.43) for U140, IRR 535,052 ($66.88) for S130, IRR 525,044 ($65.63) for S140	Iranian rials (IRR). The conversion rate to US dollars ($) was $1 = IRR 8,000	Direct costs (immediate direct costs incurred by the hospital laboratory in the screening and diagnosis of GDM)	Discount was not relevant and not reported

Author/ year/ Country and city	Perspective/ Price year	Intervention/ Comparator	Selective screening criteria for high risk women	Clinical effectiveness measures/ Resource	effectiveness outcomes	Cost measures/ Resource	Cost reference	Cost outcomes	Currency	Direct or indirect costs	Discount
Reed/ 1984 USA, Salt Lack city	Hospital perspective/ N/A	(1) Universal screening using 50g GCT (150mg/dl)+ 100g OGTT; (2) Selective screening (traditional risk factors) using 50g GCT (150mg/dl)+ 100g OGTT; (3) Selective diagnosis using 100g OGTT (4) Universal diagnosis using 100g OGTT (5) Selective screening (aged over 25) using 50g GCT (150mg/dl)+ 100g OGTT.	Criteria 1:family history of diabetes, a prior macrosomic infant (≥ 9lb), history of adverse obstetric outcome (two or more pregnancies of fetal death, neonatal death, congenital anomly, prematurity, excessive wight gain, hypertension, or proteinuria) Criteria 2: age ≥25	Number of cases missed/A precious study of O'Sullivan *et al.* (1973)	Number of cases missed: (1) 5 (20%), (2) 15 (60%), (3) 12 (48%), (4) 0 (0%), (5) 6 (24%).	Cost per GDM case detected/ A precious study of O'Sullivan *et al.* (1973)	N/A	Cost per GDM case detected: (1) $684.40, (2) $683.18; (3) $938.46, (4) $976.00; (5) $386.11	US dollars	Direct costs (immediate direct costs incurred by the hospital laboratory in the screenig and diagnosis of GDM)	Discount was not relevant and not reported

Author/ year/ Country and city	Perspective/ Price year	Intervention/ Comparator	Selective screening criteria for high risk women	Clinical effectiveness measures/ Resource	effectiveness outcomes	Cost measures/ Resource	Cost reference	Cost outcomes	Currency	Direct or indirect costs	Discount
Shamsuddin/ 2001 Malaysia	Hospital perspective/ N/A	(1) Universal screening using 75g OGTT; (2) Selective screening using 75g OGTT.	Age≥35, obesity (weight ≥80kg), family history of diabetes, personal history of GDM, a prior macrosomic infant, history of adverse obstetric outcome, urinary tract infection, vaginal discharge and pruritis vulvae, glycosuria	Sensivitity /A cross-sectional survey of 768 pregnant women	(1) Universal screening: 100% sensitivity; (2) Selective screening: 72.2% sensitivity	Cost per GDM case detected/A cross-sectional survey of 768 pregnant women	N/A	Cost per GDM case detected: (1) Universal screening: RM 12.06; (2) Selective screening: RM 11.15	Ringgit Malaysian (RM). The conversion rate to US dollars ($) was $1 = RM 3.80	Direct costs (immediate direct costs incurred by the hospital laboratory in the screenig and diagnosis of GDM)	Discount was not relevant and not reported

3.4.3.3 Data synthesis of the cost-effectiveness and cost studies

The cost-effectiveness study (Poncet *et al*., 2002) used a health system perspective. The cost measures included screening tests, obstetrical cares, management of GDM, and if any, delivery cares, and sick leave, starting from the 24th week of gestation till discharge from maternity. The outcome measures of effectiveness were macrosomia, prematurity, perinatal mortality, and hypertensive disorders rates. Poncet *et al*. (2002) found the cost to obtain one unit of additional effectiveness under universal screening was 1.1 times more expensive than selective screening, thus suggested selective screening with two-step tests was more cost-effective than universal screening. However, this study neither considered the long-term consequences of GDM after delivery nor the potential cost of missing a diagnosis of GDM through selective screening. Admittedly, these are difficult to measure, nevertheless these limitations could potentially affect the ICER of 1.1 reported in this study.

All cost studies used hospital respective considering the direct costs of GDM tests. The main outcome of cost per GDM case detected ranged from $2.9 for selective screening (while $3.3 for universal screening) in Malaysia (Shamsuddin *et al*., 2001) to $386.11 for selective screening (while $684.40 for universal screening) in the USA (Reed *et al*., 1984). The GDM incidence in the Malaysia study (24.9%) was ten times of the GDM incidence in the US study (2.5%), which partly explained the big difference in cost per case detected. However, the Malaysia study failed to report the detailed cost per unit of GDM screening, which made further comparison of the screening cost between the two studies impossible. Reed et al. (1984) found the proportions of missed GDM were similar between selective screening (24%) and universal screening with two-step tests (20%), when compared with universal screening with one-step test, while selective screening for women over 25 years old was considerably cheaper in terms of cost per GDM case detected ($386.11 vs.

$684.40), thus suggested selective screening for women aged over 25. The other three cost studies all recommended universal screening as they found the cost per GDM case detected under selective screening approach was just slightly cheaper than universal screening, however the selective approach missed a significant proportion of GDM cases (Coustan *et al.*, 1989; Larijani *et al.*, 2004; Shamsuddin *et al*., 2001).

3.5 DISCUSSION

3.5.1 Statement of principal findings

Only seven out of 28 effectiveness studies recommended selective screening (Caliskan *et al*., 2014; Helton *et al*., 1997; Naylor *et al*., 1997; Sacks *et al*., 1987; Van Leeuwen *et al*., 2010; Jensen *et al*., 2013; Pintaudi *et al*., 2014). Fifteen studies did not recommend selective screening over universal screening either because of low specificity (proportion of women who could be exempted from the screening) or low sensitivity (proportion of GDM cases identified) of the selective screening approach. Two studies recommended selective screening under certain circumstances (Jimenez-Moleon *et al*., 2002; Savona-Ventura *et al*., 2013), two studies required further research (Moses *et al*., 1998; Corrado *et al*., 2014), and the remaining two studies did not conclude with a recommendation (Shirazian *et al*., 2009; William *et al*., 1999)

The studies which recommended selective screening had been carried out in areas where GDM prevalence was relatively low compared to the contexts of studies wherein universal screening was recommended. The studies which recommended selective screening involved all four studies that used their own developed selection

criteria for identifying women at high risk (Caliskan *et al*., 2014; Naylor *et al.*, 1997; Van Leeuwen *et al*., 2010; Pintaudi *et al.,* 2014), including the use of a risk scoring algorithm (Naylor *et al.,* 1997; Van Leeuwen *et al*., 2010).

Only one of the four cost studies recommended selective screening for women over 25 years old (Reed *et al*., 1984). The other three studies favoured universal screening because the cost per GDM case detected was just slightly higher than that for selective screening, but this did guarantee that no GDM cases were missed (Coustan *et al.*, 1989; Larijani *et al.*, 2004; Shamsuddin *et al*., 2001). The cost-effectiveness study favoured selective screening with two-step tests, since the ICER of universal screening was 1.1 times that of the ICER of selective screening (Poncet *et al*., 2002).

3.5.2 Strengths, limitations, and uncertainties of the review

Four main databases of Medline, EMBASE, Web of Science, and Cochrane Database were searched, which covered key resources. A second reviewer conducted a kappa analysis of 20% of the studies at title, abstract, and full text stages for the screening process. An example data extraction result of an included study was also double-checked by the second reviewer. Disagreements were discussed and agreements were achieved. This minimised the researcher bias of occasional inappropriate performance in the process of selection and data extraction for the review.

As mentioned in section 3.4.2.1 earlier, the average quality score of the included studies was 13.4 out of 16. The reviewed studies reported excellently on aims and objectives, intervention and comparator. However, some of them failed to describe the outcome measures clearly in the introduction or methods section. When there was more than 10% loss to follow-up, some studies neither described the

characteristics of these patients nor took these into account in the final analysis. The majority of studies did not report the sample size calculation for their study population (please refer to section 3.4.2.1 for more details of the frequently failed items). These drawbacks can be seen as having possible implications for their final outcomes and conclusion.

Due to the language limit for English written articles only, some potentially valuable articles written in other languages could have been missed. One other limitation is that the review assessed the efficacy outcomes of sensitivity and specificity, whereas the clinical outcomes were not assessed. However, there was a reason for it. The outcome measures of the current review changed from clinical outcomes to efficacy outcomes (sensitivity and specificity) after conducting a scoping search. Only three studies (Griffin *et al.,* 2000; Cosson *et al.,* 2006; Ezimokhai *et al*., 2006) were identified which used the clinical outcome measures. The three studies found the clinical outcomes of the GDM women diagnosed by universal screening were better than those GDM women diagnosed by selective screening. However, these findings could not be used to conclude that universal screening was better as the researchers had done because the GDM women diagnosed by selective screening represented more severe cases as they had risk factors, thus indicating the potential for worse clinical outcomes. The most appropriate study design for a clinical outcome comparison should be to compare the clinical outcomes of all women not just GDM women under the two screening approaches. No study was found that compared the clinical outcomes of all pregnant women under the universal screening versus selective screening. Having said so, if we assume that the GDM diagnosis criteria is correct (i.e., all the women diagnosed are GDM women, and these GDM women need treatment), then exploring the efficacy outcomes is also enough to compare the two screening approaches.

3.6 CONCLUSION

3.6.1 Implications for service provision

The results of the systematic review showed that only seven of 28 effectiveness studies recommended selective screening. These seven studies were conducted in areas of relatively low GDM prevalence, and included all four studies which used their own selection criteria for selecting high risk women rather than using the standard guideline criteria. For areas with high GDM prevalence, universal screening for GDM is recommended unless further evidence emerges. For areas with low GDM prevalence, it is recommended that on the one hand, they continue their current practice (whether it is universal or selective screening) before any robust up-to-date evidence fight against the current practice (the evidence should share the same setting with them including same country, same selection criteria, and same OGTT method); on the other hand, they are encouraged to explore the evidence themselves regarding the effectiveness of a selective screening approach. There is no definitive cut-off for judging whether a country's GDM incidence is low or high. The NICE guideline suggested certain high-risk ethnical groups including South Asia (India, Pakistan, Bangladesh), Black Caribbean and Middle Eastern origin (Saudi Arabia, United Arab Emirates, Iraq, Jordan, Syria, Oman, Qatar, Kuwait, Lebanon, Egypt) (Webber *et al*., 2015) for GDM screening. From the current systematic review, a cut-off of 3.2% (this is the mean GDM incidence of the five studies which recommended selective screening as indicated in page 66) under two-step GDM screening and a cut-off of 6.9% (this is the mean GDM incidence of the two studies which recommended selective screening as indicated in page 69) under one-step GDM screening might be considered.

From an economic perspective, the cost-effectiveness study and three of the four cost studies suggested that universal screening was only slightly more expensive than selective screening. Unless further evidence arises, the economics is of less concern when deciding which screening approach to use.

3.6.2 Suggested research priorities

Only four studies explored the effectiveness of selective screening using their own developed selection criteria, based on the local population (Caliskan *et al*., 2014; Naylor *et al.*, 1997; Van Leeuwen *et al*., 2010; Pintaudi *et al.*, 2014), all of which found selective screening were effective and recommended selective screening. Among them, two studies developed and used a risk scoring algorithm (Naylor *et al.*, 1997; Van Leeuwen *et al*., 2010). It is recommended that more studies be conducted to assess the effectiveness of selective screening using a self developed risk criteria which is specific to the local population characteristics, especially for areas with relatively low GDM prevalence. Furthermore, more cost-effectiveness studies using the perspective of health system are needed, since the evidence of one cost-effectiveness study is not enough to make a definite conclusion.

The new IADPSG criteria was recommended by the WHO (WHO, 2013) and was increasingly adopted by countries worldwide including China, Italy and Australia (Chinese Ministry of Health, 2011; Wong, 2014). However, this review found no study from these countries comparing universal versus selective screening under the new IADPSG OGTT criteria, further studies in this area were therefore suggested. Chapter 5 of this PhD research evaluated the effectiveness of a risk score-based selective screening approach compared with universal screening under the IADPSG criteria in China. Further discussion in the context of the findings from chapter 5 were made and included in the Overall Discussion chapter at the end of the thesis.

When conducting future comparative studies on universal versus selective screening, there are several suggestions for a high quality study. Firstly, describe the outcome measures clearly in the introduction, method, and result sections; a clear definition and presentation of sensitivity and specificity should be reported. Secondly, make sure that the study population is representative of the entire pregnant women population, justify and clearly describe the inclusion and exclusion criteria. Thirdly, report the refusal rate and loss to follow-up rate where applicable, if these rates exceed 10%, take them into account in the final analysis. Fourthly, remember to report the sample calculation which was frequently ignored by previous studies. Finally, if the clinical outcomes are to be assessed, make sure that the clinical outcomes of all pregnant women under the universal screening group versus selective screening group are compared. Undertaking the above recommended research will fill the current evidence gap and generate more useful knowledge, contributing to a more robust conclusion about universal versus selective screening.

The findings of the current systematic review help to address the overall thesis aim of exploring the most appropriate screening approach for GDM and, in particular, comparing the effectiveness and cost-effectiveness of universal versus selective screening. Further considerations with reference to the overall thesis aim will be presented in Chapter 4 and Chapter 5.

Chapter 4: Pregnant women's attitudes, views, and experience of the IADPSG universal screening approach for GDM in China: a Q methodology study

4.1 BACKGROUND

4.1.1 GDM background

GDM is defined as any degree of glucose intolerance that occurs or is first recognised during pregnancy (Turok *et al.*, 2003). Incidence of GDM ranges from less than 1% to 28% worldwide (Jiwani *et al.*, 2012). In China, GDM incidence was between 8% to 15% under the IADPSG criteria (Shang & Ma, 2011; Wei & Yang, 2011; Hou *et al.*, 2012; Lu *et al.*, 2012; Jiang *et al.*, 2013). GDM can lead to various adverse maternal and fetal outcomes, including spontaneous abortion, caesarean section delivery, and pre-eclampsia to the mother; stillbirth, macrosomia, and shoulder dystocia to the newborn (Odar *et al.*, 2004). Both mother and newborn are at increased risk of developing type 2 diabetes in the future (Damm *et al.*, 2009).

4.1.2 The IADPSG approach for GDM

There is no clear consensus on the best screening approach for GDM. As explained in section 1.3.1 in Chapter 1, having all pregnant women (universal screening) or only high risk pregnant women (selective screening) to undergo the GDM tests, undergoing two-step GDM tests or a one-step GDM test, are all different possible options. Before 2011, China was implementing two-step universal screening approach for GDM. All pregnant women in China underwent a 50g glucose challenge test (GCT) during the 24^{th} to 28^{th} weeks of gestation; those with abnormal values then further undergo a 75g or 100g OGTT for diagnosis. Diagnosis was confirmed with at least two abnormal values from four measured: blood glucose ≥5.8mmol/l at fasting, 10.6mmol/l at 1 hour, 9.2mmol/l at 2 hours, and 8.1 mmol/l at 3 hours (Yang, 2009). With a 50g GCT beforehand, only a small number of women need to further undergo the complicated OGTT diagnostic test.

In 2010, the IADPSG suggested a one-step universal screening approach for GDM (Panel, 2010). All pregnant women, unless already diagnosed with diabetes, undergo the OGTT directly during the 24^{th} to 28^{th} weeks of gestation. Diagnosis of GDM is made if there is at least one abnormal value from three measured: blood glucose ≥5.1 mmol/l at fasting, 10mmol/l at 1 hour, and 8.5mmol/l at 2 hours. All diagnosis thresholds of OGTT used in this new approach were lower than previous standards.

In July 2011, the Chinese Ministry of Health (MOH) adopted the IADPSG one-step universal approach for GDM (Chinese Ministry of Health, 2011), which is expected to bring both advantages and disadvantages to pregnant women in China. One possible advantage is that women who might have adverse obstetric and perinatal outcomes due to higher glycemic levels are less likely to be missed for GDM treatment. Another advantage is that GDM women (the minority) are now diagnosed with GDM in just one visit, as opposed to two (1-hour 50g GCT screening test followed by 3-hour OGTT diagnosis test). The disadvantage of the IADPSG approach is that the non-GDM pregnant women (the majority), who previously only needed a relatively simple 50g GCT screening test (non-fasting, one hour, one blood sample required) now have to undergo the more complicated OGTT test (fasting, two hours, three blood samples required). GDM incidence is estimated to increase by two to three fold using the IADPSG approach, but it is not completely confirmed that these additionally identified women will benefit from treatment, or to what extent (Vandorsten *et al.*, 2012).

4.1.3 User perspectives on the IADPSG approach for GDM

As described in section 1.3.5 in Chapter 1, patients are one of the key stakeholders whose perspectives and needs should be heard and integrated in any healthcare decision making and evaluation. It is important to understand pregnant women's attitudes and views as well as their experiences of the one-step GDM diagnosis

approach. Griffiths *et al.* (1993) explored the attitudes of pregnant women towards universal one-step GDM screening with modified OGTT (two blood tests at fasting state and after 1 hour; a 75g glucose load is taken either at home or in a collection center) in Australia, with generally positive results. The Australian study was positive about the convenience of the screening method used, the need for all women to be screened in pregnancy, and their desire to be screened in future pregnancies.

However, the study was conducted more than 20 years ago, and no study has been conducted recently to understand women's attitudes or experiences of GDM screening, especially in the case of the new IADPSG universal one-step screening approach (three blood tests at fasting state, after 1 hour, and 2 hours; a 75g glucose load is usually taken at the hospital) for GDM. There is thus a gap in knowledge in regards to pregnant women's views of the GDM test, the GDM information they received before and after the test, and their feelings towards the screening process and result. To fill the gap, this study used a Q methodology design to explore women's attitudes, views, and experience of the IADPSG universal screening approach for GDM in China.

4.2 AIM AND OBJECTIVES

The study aimed to investigate pregnant women's attitudes, views, and experience of the IADPSG one-step universal screening approach for GDM in China.

The specific objectives were:

(1) to develop a list of Q statements for assessing pregnant women's attitudes, views, and experience of the one-step GDM diagnosis approach;

(2) to implement the Q methodology study and collect data from pregnant women, comprising 15 diagnosed with and 15 without GDM;
(3) to analyse pregnant women's responses to provide user perspectives concerning the IADPSG recommendations based one-step universal approach for GDM in China.

4.3 METHODOLOGY

4.3.1 Q methodology

The Q methodology is a systematic study of subjectivity, i.e., a person's viewpoint, opinion, beliefs and attitudes (Brown, 1993). Q methodology was originally invented by British psychologist William Stephenson in 1935. The use of Q methodology has expanded to other fields outside psychology, most notably in the area of political, social, and health sciences (Brown, 1997). In a Q methodology study, participants are provided with a set of statements on a certain topic, called the Q-set. Respondents, called the P-set, are asked to rank-order the statements into a subjectively meaningful pattern (Q-sort) from their individual point of view, according to their feelings, preferences, and judgments (Van Exel & de Graaf, 2005). Resulting Q-sorts are analysed using correlation and factor analysis (Q-analysis), yielding a set of factors whose interpretation reveals a set of points-of-view (the F-set) (Watts & Stenner, 2005).

The aim of a Q methodology study is to establish the existence of distinct viewpoints and thereafter to understand, explicate, and compare findings (Brown, 1980). To achieve this, engagement of very few individuals or even a single person could be appropriate (Watts & Stenner, 2005). Although 40-60 participants are considered suitable in the UK tradition of multiple-participant Q methodology (Stainton Rogers,

1995), good analyses might be conducted with considerably less (Stephenson 1953; Watts & Stenner, 2005). Q methodology suggests a minimum ratio of two Q-set items to every participant, in practice, it is sensible to choose a number of participants that is less than the number of Q-set items (Watts & Stenner, 2005).

Q methodology utilises by-person factor analysis, rather than the traditional by-variable analysis. The Q-set items are treated as sample while the participants are variables (Watts & Stenner, 2005). The analysis aims to establish the diversity and range of viewpoints expressed by the participants (Cross, 2005). The proportions of individuals in a factor are not revealed in the factor analysis, but the distinctive points of view in the form of statements that distinguish each other are revealed (Ward, 2010). Q sorting uses a quasi-normal distribution (Van Exel & de Graaf, 2005); an example of Q sorting from 'agree least' to 'agree most' is shown in Appendix 6. Although Q methodology analysis could be achieved using IBM SPSS statistics, the ideal analysis software package to use is PQMethod, which has been purpose-built to do Q analysis and is free to download (Watts & Stenner, 2005).

4.3.2 Rational of choosing Q methodology

There are both qualitative and quantitative techniques to explore the perceptions, attitudes, or viewpoints regarding a research question. Both techniques have strengths and weaknesses. Brown (1996) criticised the problems with the qualitative/quantitative dichotomy in research, and promoted Q methodology which was considered to combine the strengths of both qualitative and quantitative research. Wilson (2005) described Q methodology as a technique that has the same level of mathematical rigor as quantitative methodology, and meanwhile has an interpretive component comparable to that of qualitative methodology. Cross (2005) stated that Q methodology allowed for the simultaneous study of objective and subjective issues to determine the perceptions.

Q methodology is appropriate for the current study, which is about exploring pregnant women's perspectives regarding GDM screening. Use of Q methodology overcomes the drawbacks of traditional survey or interview. Traditional survey imposes meaning a priori and investigates pre-specified categories. However in Q methodology, it allows individuals to determine what is meaningful, valuable, and significant from their point of view (Ward, 2010). Interviews can be intrusive, time consuming, and sometimes inefficient. The current study involves pregnant women who are in vulnerable physical and potentially psychological condition. A Q methodology study would be more efficient in keeping the study focused and avoiding these risks.

Since Q methodology uses a small sample size to investigate human subjectivity, the reliability or generalisability of the study results have often been criticised (Thomas & Baas, 1992). According to Brown (1980), an important notion behind Q methodology is that only a limited number of distinct viewpoints exist on any topic. Any well-developed Q sample, containing the wide range of existing opinions on the topic, will reveal these viewpoints (Van Exel & de Graaf, 2005). Thomas and Baas (1992) concluded that the reliability and the ability to generalise the sample results of Q are of less concern. The Q methodology results are the distinct subjectivities about a topic that are operant, not the percentage of the sample (or the general population) that adheres to any of them (Van Exel & de Graaf, 2005).

4.3.3 Development of the Q statements

In the current study, a concourse of 90 statements was initially developed regarding pregnant women's attitudes, views, and experience of the GDM diagnosis approach. These statements were produced based on literature review, various related websites where women express their experience and opinions (e.g., pregnant women's internet forum http://www.netmums.com and Facebook), and by consulting five new

mothers in the UK. These five mothers (one was diagnosed with GDM) who delivered within the last 6 months at the George Eliot Hospital in Nuneaton in the UK were consulted via the PRiDE event 'Little Elves party' on 10th December 2013. The researcher was assisted by the senior members of the team (PRiDE: http://www2.warwick.ac.uk/fac/med/research/mvhealth/patient_care/pride/) during the process. Then, a total of 32 Q-set items (i.e., Q statements) were established from the concourse by consulting three new mothers in China and three academic staff at the Warwick Medical School of the University of Warwick. These three new mothers (one was diagnosed with GDM) who delivered within the last year in hospitals in Chengdu in China were consulted via telephone from the researcher's personal networks in December 2013. The final 32 statements ask pregnant women's perspectives about the GDM diagnosis test, the GDM information they received and would like to receive before and after diagnosis, and their general viewpoints on the GDM diagnosis process. A full list of the 32 Q statements is shown in Appendix 7.

4.3.4 FlashQ software for Q methodology study

The use of Q methodology has been facilitated by recent developments in computer software. FlashQ is a free Flash application developed by Christian Hackert and Gernot Braehler for performing Q-sorts on a computer or online (FlashQ: http://www.hackert.biz/flashq/home/). It facilities data collection of the Q methodology study and also makes the responses much easier for participants.

The final 32 Q statements were set up using FlashQ. The screenshots illustration of how FlashQ programme works on a computer is shown in Appendix 8, including an example of the Q-sort results. The FlashQ programme has been translated into Chinese, to be conducted among pregnant women in China. The translation was done by the researcher whose nationality is Chinese, and was checked by another Chinese with a master's degree in public health.

4.3.5 Implementation of the Q methodology study in China

The Q methodology study using FlashQ was conducted with 30 pregnant women at the Chengdu First People's Hospital in China, including 15 diagnosed with and 15 without GDM.

Pregnant women visit the Department of Obstetrics for their first antenatal care at around 12th week of gestation (3 months); a record card and a paper record are established for them at this first visit. They visit the hospital every 4 weeks before 20th week of gestation, every 2 weeks between 20th and 32rd week of gestation, and every week after 32rd week of gestation. For antenatal visits before 32rd week of gestation, every pregnant woman visits the Registration Room at the first place to collect their paper record, and return back the paper record to the Registration Room after she visits her doctor.

The recruitment process took place at the Registration Room in the Department of Obstetrics. The recruitment was facilitated by the senior staff at the Registration Room. The senior staff identified the pregnant women for their gestational week and GDM status from their paper records. Pregnant women between 24th to 30th week of gestation were considered, which meant they had underwent the OGTT and received the GDM diagnosis result. The staff asked the eligible women whether they are interested to take part in a small study when they collected from or returned their paper records to the senior staff, and referred the women to the researcher who sat opposite to the staff at the Registration Room. Interested pregnant women were given the 'Participant Information Leaflet' by the researcher and were given time to read the leaflet in the waiting area beside the Registration Room. Women who decided to participate were led to a quiet Outpatient Room 10 metres away from the Registration Room by researcher to undertake the study. A 'Consent Form' was

given to the participant by the researcher to sign, and was collected before the study started. All women were given the opportunity to ask the researcher any questions about the study before signing the consent forms. The study was undertaken on a laptop computer in the Outpatient Room. An instruction sheet of FlashQ was given to the participants for guidance (see Appendix 9). The first 15 GDM women and the first 15 non-GDM women who consented to take part in the study were recruited. A small souvenir from the University of Warwick was given to each woman who participated in the study.

4.3.6 Outcome analysis

The analysis of the Q sorts is an objective computerised process, which is one of the scientific base of Q methodology (Van Exel & de Graaf, 2005). PQMethod (version 2.11) was used for Q sorts analysis (PQMethod: http://schmolck.userweb.mwn.de/qmethod/#PQMethod), which is a free statistical programme tailored for Q methodology analysis. Its original fortran programme, QMethod, was developed by John Atkinson at Kent State University in 1992, and was ported by the maintainer of the programme site to the PC and updated with added features to versions later on. Brown (1980; 1993) provided a comprehensive overview of the analysis of the Q sorts. First, the correlation matrix of all Q sorts is calculated, which reveals the degree of (dis)similarity in points of views between the individuals. Second, the correlation matrix is subject to factor analysis, which examines how many basically different Q sorts are in evidence. Third, the original set of factors is then rotated to arrive at a final set of factors. Finally, the calculation of factor scores and difference scores are done, which points out the salient statements that deserve special attention in describing and interpreting that factor. Interpretation of the results is based on these factor scores and difference scores. The explanations individuals give following their ranking are used to interpret the factors and as illustration material.

The qualitative comments given by Chinese participants were translated into English for interpretation.

4.3.7 Ethical considerations

Pregnant women were fully informed of the study; the 'Participant Information Leaflet' and 'Consent Form' were distributed to pregnant women, and they were free to decide whether or not to participate. The information leaflet also clarified their right to withdraw from the study freely at any time. If a woman liked to participate, her consent form was completed and signed before any data are collected. Any woman who withdrew after providing consent to take part in the study were recorded and replaced, to ensure 15 GDM and 15 non-GDM participants.

Neither names nor any other identifying information of participants were collected. Each woman was anonymised and given a serial number as a unique ID. Collected study information, including the Q methodology study results and consent forms, could only be accessed by the researcher and the research team from the University of Warwick, and the collaborators from the Chengdu First People's Hospital. After the study, the electronic FlashQ results were stored securely in a computer with access password in the office at the Warwick Medical School, the University of Warwick, while paper records of the consent forms were stored in a locked cabinet in the same office.

Ethical consent to conduct this study was sought initially in China, where the Hospital Ethics Approval Committee of the Chengdu First People's Hospital scrutinised and approved the research proposal. A further University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC) approval to conduct this research as part of a Health Sciences PhD programme was sought and granted.

The BSREC reference is REGO-2014-704. Please see the full approval letters in Appendix 10.

4.4 RESULTS

4.4.1 Data collection results

4.4.1.1 Recruitment and data completion

Please see Appendix 7 for a list of 32 developed Q statements, and Appendix 8 for the established FlashQ programme for the Q methodology study. Initially, 17 GDM women and 16 non-GDM were recruited to undertake the study. Three were excluded since they withdrew in the middle and failed to complete the whole study. 15 GDM and 15 non-GDM women who fully completed the study were finally involved in the study. The FlashQ software automatically generated the result sheets in XPS document format after they finished the study.

The participants were required to provide reasons for the two Q statements they most agreed with and the two statements they most disagreed with after all the statements had been rank ordered. To facilitate this process and to be consistent with all participants, the researcher typed their reasons into the FlashQ software.

4.4.1.2 Basic hospital information related to the study

The Chengdu First People's Hospital is a top-class hospital (tertiary referral hospital) in Chengdu, the capital city of Sichuan Province in China. The Department of Obstetrics and Gynaecology at the hospital serves more than 3000 deliveries each

year. Each pregnant woman registered at the Registration Room in the Department of Obstetrics and Gynaecology of the hospital when they first visited the Department during 12th to 14th gestational week. The OGTT takes place during the 24th to 28th gestational week.

At the bottom of the pregnant women's registration card is brief information about the OGTT test. No additional leaflet or material of GDM/OGTT was provided to pregnant women. No bulletin was available on GDM/OGTT on the walls in the Department of Obstetrics and Gynaecology. There was a "pregnant women's school" at the hospital which provided free weekly courses relevant to pregnancy. However, no GDM/OGTT course was provided. The weekly school is non-obligatory, and attendance is middling.

There was a scanning machine for report collection, which was located on the same floor as the Department of Obstetrics and Gynaecology. Pregnant women wait several hours after OGTT to collect their OGTT result from the scanning machine. They then give the results sheet to their doctor so that they can receive explanation and advice. If a pregnant woman is diagnosed as having GDM, she will then be referred to the Department of endocrinology at the hospital for further advice and treatment.

4.4.2 Data analysis result

4.4.2.1 The analysis process

The analysis followed the standard procedure using the PQMethod software (Figure 11), under the instruction of the PQMethod Manual (http://schmolck.org/qmethod/pqmanual.htm).

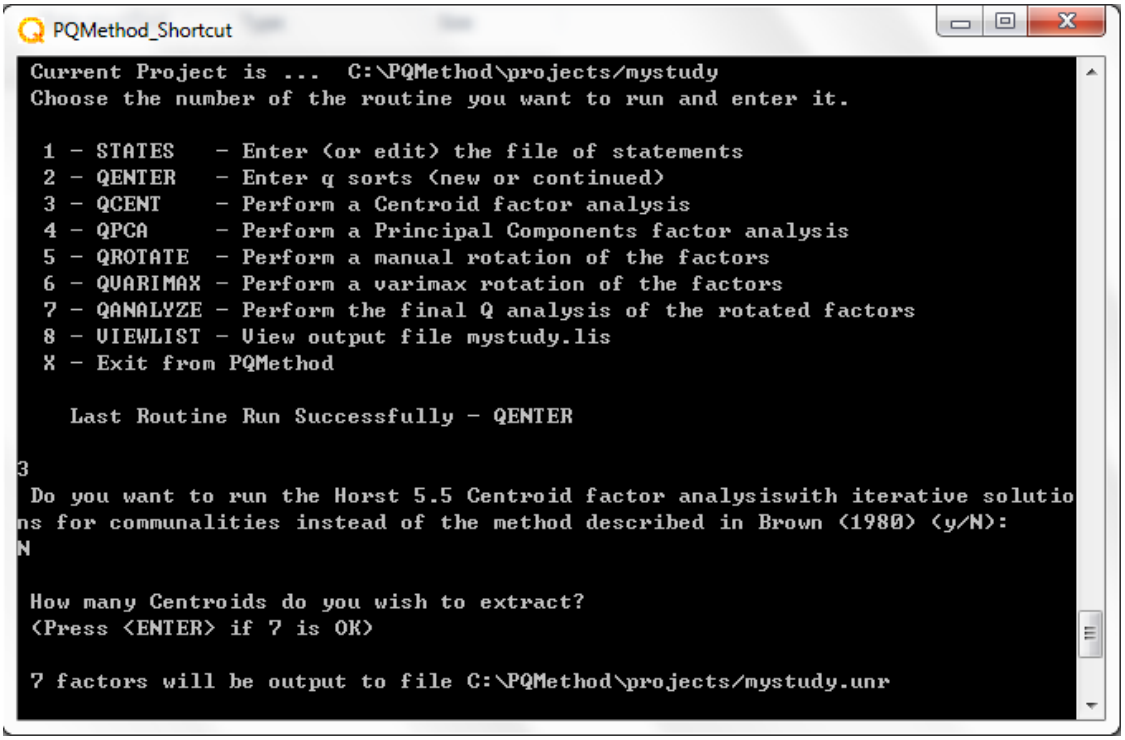


Figure 11. Standard analysis procedure in the PQMethod software

First, '1-STATES' was performed to enter the 32 Q statements. Second, '2-QENTER' was performed to enter the Q sort of each of the 30 participants. Q sorts of the 15 non-GDM women were entered first, followed by Q sorts of the 15 GDM women.

'3-QCENT' (Centroid analysis) and '4-PCA' (Principal Component analysis)' were two alternative options for extracting (unrotated) factors. The QCENT offers the option of choosing between two methods of Centroid extraction, the customary method described in Brown (1980) and the Horst's (1965) method. The Brown Centroids, as a default, suggests the 'magical number' of seven centroids to extract (centroids are factors). '5-QROTATE' and '6-QVARIMAX' are alternative options for rotating the factors. '5-QROTATE' allows rotation by hand, whereas '6-QVARIMAX' is automatic rotation (PQMethod Manual - version 2.35)

As pointed out by Watts & Stenner (2005), there was often little reason for preferring one method over another for factor extraction and factor rotation. Though centroid analysis is generally preferred, the PCA analysis is equally satisfying. For factor rotation, in practice many Q methodologists prefer the simplicity and reliability of the varimax procedure, which automatically seeks the mathematically superior solution (Watts & Stenner, 2005). I chose 3-QCENT and 6-QVARIMAX since these two methods were more standardised and automated, and reduced the researcher bias of making different selections when using 4-QPCA and 5-QROTATE.

Finally, '7-QANALYZE' was conducted to perform the final Q analysis of the rotated factors. '8-VIEWLIST' was used to view the output file in '.lis' format, which could be opened with Notepad.

4.4.2.2 Factor extraction result

Having conducted '3-QCENT' (Centroid analysis) using the 'Brown Centroids' method, seven factors were extracted. The end product of the factor extraction process is a table of factor loadings indicating the initial association, or correlation, of each Q sort with each factor (Table 10). The table that is produced is also called an unrotated factor matrix. The factor loading (also known as factor saturation) is expressed in the form of a correlation coefficient. It tells us the extent to which each individual Q sort can be said to exemplify, or is typical of a factor pattern (Watts & Stenner, 2012).

Table 10. The Un-rotated Factor Matrix Generated by the Centroid Factor Analysis

		Factors						
	Sorts	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	P004	0.6088	-0.2786	-0.2632	-0.0098	-0.4333	0.0412	0.1526
2	P005	0.7363	0.1622	-0.1214	0.068	0.137	0.0144	0.0106
3	P006	0.542	-0.3372	0.21	-0.3585	-0.2399	0.0612	0.1254
4	P007	0.5357	0.396	0.2472	0.0776	0.1194	0.0876	0.0303
5	P008	0.6216	-0.3305	0.1686	-0.0682	-0.2673	0.0587	0.0634
6	P013	0.5627	0.4495	-0.2189	-0.1956	-0.2086	0.1145	0.0743
7	P014	0.6325	0.1883	-0.1458	0.0856	-0.2298	0.0193	0.0511
8	P015	0.7743	0.3031	0.3454	0.0804	0.1142	0.0506	0.0576
9	P016	0.5204	0.5202	0.2445	0.2256	-0.297	0.1568	0.1112
10	P017	0.3189	-0.2663	-0.1286	-0.4792	0.1522	0.0376	0.124
11	P019	0.8341	0.1064	0.0829	-0.0522	-0.0131	0.0065	0.005
12	P020	0.1143	-0.5378	0.2914	0.0401	0.0099	0.1659	0.0387
13	P022	0.3128	-0.6606	-0.407	-0.1481	0.0658	0.2657	0.0905
14	P023	0.2529	-0.4503	-0.3019	0.4777	0.0699	0.1127	0.1564
15	P027	0.5429	-0.2722	0.5842	0.0846	0.0785	0.0393	0.169
16	PG001	0.4242	-0.1486	0.2203	-0.3009	0.2991	0.0114	0.0927
17	PG009	0.5835	-0.0551	-0.4135	0.3873	-0.0951	0.0015	0.1705
18	PG010	0.2486	-0.0274	0.1877	-0.0528	-0.1873	0.0003	0.0437
19	PG011	0.4864	0.4097	-0.2558	0.1322	0.0316	0.0941	0.0382
20	PG012	0.2495	0.169	-0.0497	-0.3003	0.0419	0.0156	0.0427
21	PG018	0.4748	-0.3683	-0.1257	0.115	0.2723	0.0736	0.0337
22	PG021	0.15	-0.1926	-0.3296	-0.1222	0.3412	0.0193	0.098
23	PG024	0.5231	-0.1518	0.2429	0.324	0.2519	0.0119	0.0934
24	PG025	0.5925	-0.3197	0.0539	-0.122	0.0242	0.0548	0.0072
25	PG026	0.6883	0.2395	0.2707	-0.1755	0.1149	0.0313	0.0481
26	PG028	0.6911	-0.0343	0.0054	0.1779	-0.2189	0.0006	0.0488
27	PG029	0.3538	0.2027	-0.1965	0.1481	-0.0352	0.0224	0.0311
28	PG030	0.7141	0.3191	-0.3317	-0.2776	-0.1886	0.0562	0.1207
29	PG032	0.5907	0.4173	-0.0424	0.0988	0.132	0.0978	0.0063
30	PG033	0.2997	0.5439	0.1856	0.1445	0.2733	0.173	0.0451
	Eigenvalues	8.5127	3.3601	1.9397	1.4292	1.1671	0.2335	0.2306
	% expl.Var.	28	11	6	5	4	1	1

The Eignvalue (EV) in the table (Table 10) is indicative of a factor's statistical strength and explanatory power. The absolute and relative sizes of the Eignvalues of the factors is a major determinant on how many factors to keep for following rotation. One standard requirement is to select only those factors with an eigenvalue

greater than 1.00 (Watts & Stenner, 2005), so called Kaiser-Guttman criterion (Guttman, 1954; Kaser, 1960, 1970). This cut-off was used because an extracted factor with an EV of less than 1.00 actually accounts for less study variance than a single Q sort (Watts & Stenner, 2012). Table 10 shows that Factor 1-5 has eigenvalues in excess of 1.00.

Furthermore, the Cattell's (1966) test prevents the arbitrary retention of all factors with EVs greater than 1.00 (Watts & Stenner, 2012). This method (Cattell, 1966) is more cautious than the Kaiser-Guttman criterion (Guttman, 1954; Kaser, 1960, 1970). This scree test involves the plotting of these EVs on a line graph, the number of factors to extract for rotation is indicated by the point at which the line changes slope. A scree test figure was drawn (Figure 12).

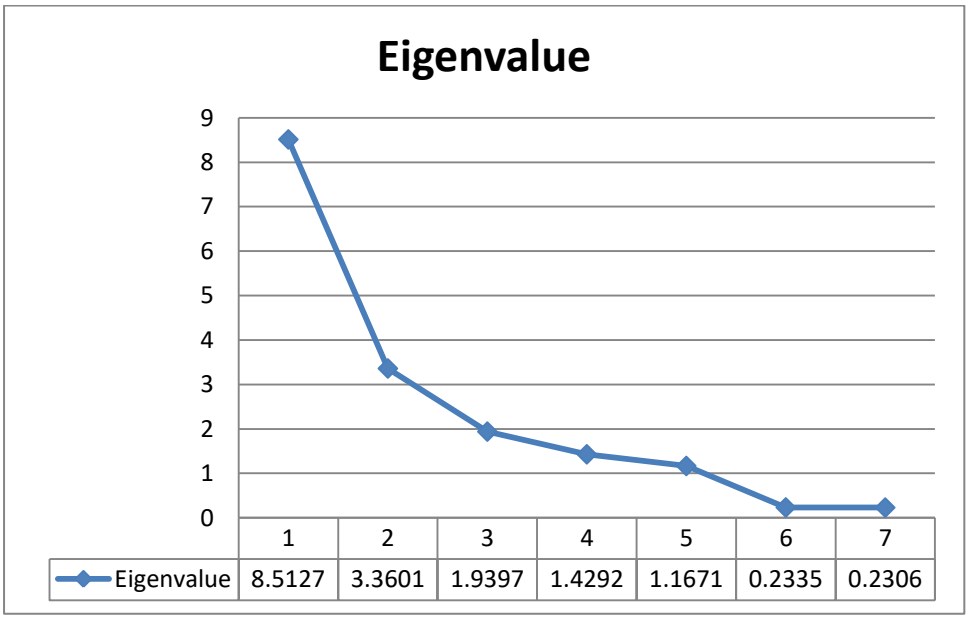


Figure 12. Scree test of seven principle components

Inspection of Figure 12 shows the slope changes from 114 ° at factor 1 to 148 ° at factor 2 with 34 ° increase. The slope then changes from 148 ° at factor 2 to 165 ° at factor 3 with 17 ° increase. The point at which the line changes slope is at factor 2.

This indicates that only two factors should be extracted from the dataset for following factor rotation and flagging.

4.4.2.3 Factor rotation result

'6-QVARIMAX' was performed for rotation of the factors. Two factors (as indicated by the scree test) were automatically rotated. A separate MS-DOS programme 'PQROT.exe' attached to PQMethod was automatically launched for graphical factor rotation and factor flagging (Figure 13).

Flagging factors seeks to associate particular subjects with factors, which is necessary for input to the '7-QANALYZE'.

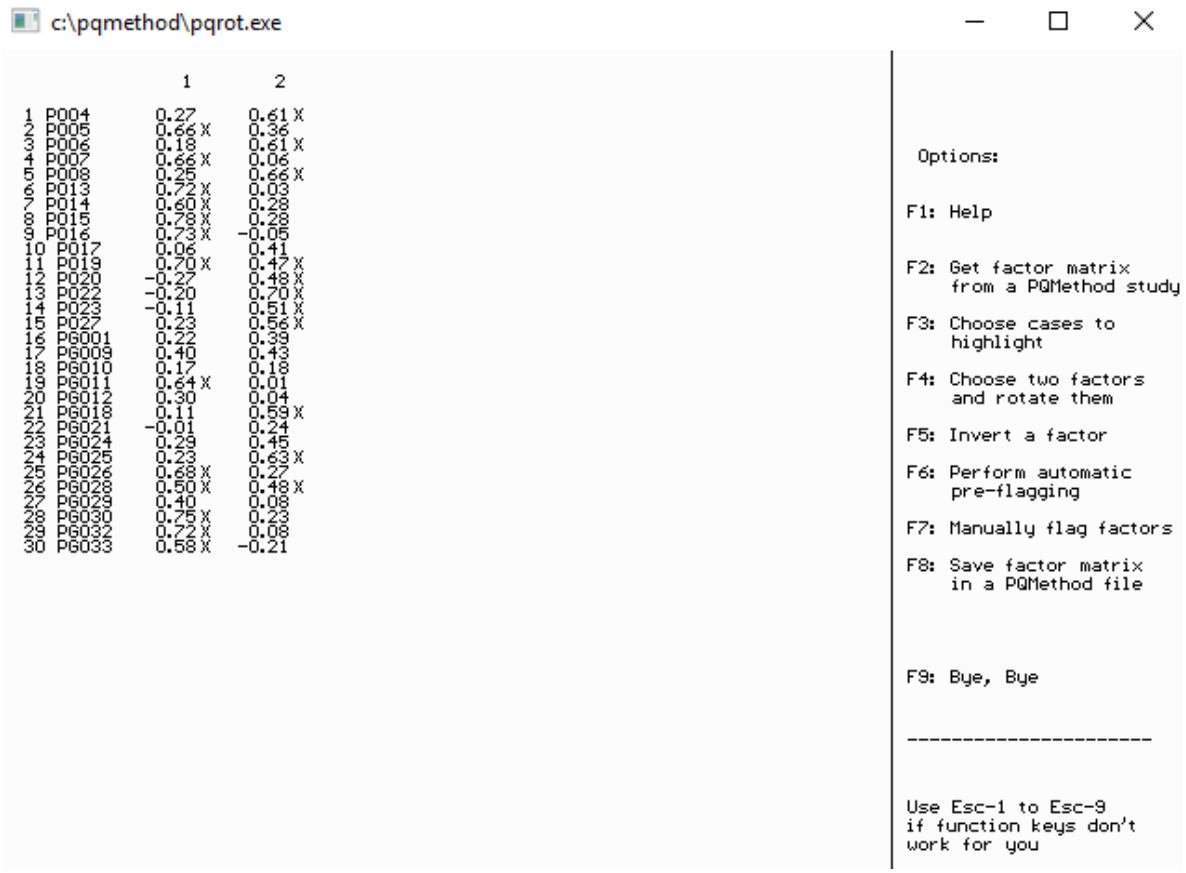


Figure 13. Flagging result for rotated factors

The factor loadings were flagged with an ‘x’ if they were a significant factor defining variants. Stephen's Table ‘quick lookup table - is it significant?’ (Appendix 11) was used for flagging based on the number of items in the study. According to the table, if a study has 32 items (Q statements), then flag anything >0.456 for a p-value < 0.01 and >0.346 for a p-value <0.05. Factor loadings >0.456 (P<0.001) was flagged for this study.

4.4.2.4 Final factor scores result

‘7-QANALYZE’ was conducted to perform the final Q analysis of the rotated factors. Only factors that were flagged would be output into the final file (.lis). The unflagged Factor 3 was dropped. The ‘8-VIEWLIST’ was used to view the output file.lis.

In effect, the Q sorts of all participants that load significantly on a given factor are merged together to yield a single (factor exemplifying) Q sort which serves as an interpretable ‘best-estimate’ of the pattern or item configuration which characterises that factor (Watts & Stenner, 2005). The factor scores (also known as factor arrays) which exemplify the best-estimate Q sorts of the two flagged factors (the factor 1 pattern and the factor 2 pattern) were displayed in Table 11 and subject to interpretation.

Table 11. Factor scores table for Factor 1 and Factor 2

Statement factors	1	2
1 The OGTT is convenient	1	-3
2 Having blood drawn three times within 2 hours OGTT is too much	-1	3
3 The duration of 2 hours for the OGTT test is too long	-1	3
4 The time from having the OGTT test to collecting the result was satisfactory	1	-2
5 I feel it is more convenient to collect my GDM diagnosis result during my next visit to the hospital, rather than waiting for several hours to collect it on that day	-1	1
6 I prefer face-to-face notification of the GDM diagnosis result when I come to the hospital rather than being telephoned when the result comes out	0	-1
7 It was easy to find the scan machine for collecting my GDM diagnosis result	2	2
8 The GDM diagnosis result sheet was NOT understandable	-2	-1
9 I was satisfied with the doctor's explanation when I gave my GDM diagnosis result sheet to him/her to review	1	0
10 I wish I had been given more information about GDM and GDM diagnosis before I had the OGTT	2	2
11 I wish I had been given more information about GDM and GDM diagnosis after I had the OGTT	2	2
12 I am NOT satisfied with the GDM information I received from the doctors/nurses at the hospital	-1	-2
13 I am satisfied with the GDM information I received from the weekly lectures for pregnant women at the hospital	0	-1
14 I am NOT satisfied with the GDM information I received from the leaflet/bulletin board at the hospital	0	0
15 I would prefer to receive GDM information from weekly lectures for pregnant women at the hospital	0	-1
16 I would prefer to receive GDM information from an education leaflet provided by the hospital	0	0
17 I would NOT prefer to receive GDM information from the hospital bulletin board	-1	-2
18 I would NOT prefer to search for GDM information (from internet, TV, magazines, books, etc.) by myself	-2	0
19 I would prefer my family (parents, husband/partner) or my friends to search for GDM information for me	-1	0
20 I would prefer to learn about GDM from talking to women who have or had GDM	0	0
21 Testing for GDM is very important and necessary	3	1
22 The OGTT should be conducted for all pregnant women	3	2
23 I feel it is a burden to undergo the OGTT	-3	-1
24 I was confused about the OGTT	-2	-2
25 I was unhappy with the OGTT	-3	-1
26 I am satisfied with the whole process of being tested for GDM	1	0
27 I am satisfied with the support I received throughout the OGTT procedure	2	1
28 I was treated with dignity and respect by staff in the hospital throughout the whole OGTT procedure	1	1
29 The doctor who gave the OGTT to me was NOT knowledgeable and informative	0	0
30 The doctor who did the OGTT listened to my concerns and listened to what I needed	0	1
31 Sometimes I felt afraid to ask the hospital staff relevant advice about OGTT and GDM	-2	-3
32 All necessary explanations and advice about OGTT were given by hospital staff	1	1

4.4.3 Interpretation of the results

4.4.3.1 Basic characteristics of participants

Thirty patients (15 non-GDM and 15 GDM women) participated in the Q methodology study. The mean age was 28.6 years old. The participants underwent the OGTT at around 25th week of gestation (ranged from 23rd to 29th week). They were interviewed about four weeks after their OGTT, which was at 29th week of gestation (ranged from 25th to 32nd week). 26.7% of the participants had not received any information about GDM before their OGTT, the number reduced to 6.6% after OGTT and before referral to the endocrinology department. 83.3% of the participants felt the Q methodology study was easy to complete (Table 12).

Table 12. Basic characteristics of participants

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
1	P004	Non-GDM	27	27	23	No	N/A	Yes	Friend	Yes
2	P005	Non-GDM	33	26	25	Yes	Friend	Yes	hospital	Yes
3	P006	Non-GDM	26	28	24	Yes	Media, friend	Yes	Media, hospital, friend	Yes
4	P007	Non-GDM	24	28	27	Yes	Hospital	Yes	Other pregnant women at hospital	Yes
5	P008	Non-GDM	25	30	24	Yes	Media, hospital	Yes	hospital	Yes
6	P013	Non-GDM	28	28	24	Yes	Hospital, friend	Yes	Hospital, friend, other pregant women at hospital	Yes
7	P014	Non-GDM	27	27	24	Yes	Media	Yes	Media, hospital	Yes
8	P015	Non-GDM	22	29	25	Yes	Media, other pregnant women at hospital	Yes	Hospital, husband	Yes

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
9	P016	Non-GDM	32	30	23	Yes	Media, hospital, friend	Yes	Media, hospital	Yes
10	P017	Non-GDM	28	30	26	Yes	Media	Yes	Media	Yes
11	P019	Non-GDM	32	28	24	Yes	Media, hospital, parents	Yes	Media,hospital, parents	Yes
12	P020	Non-GDM	31	30	27	Yes	Friend	Yes	Hospital	Yes
13	P022	Non-GDM	28	28	24	No	N/A	No	N/A	Yes
14	P023	Non-GDM	29	29	24	No	N/A	Yes	Hospital	Yes
15	P027	Non-GDM	29	27	24	Yes	Media, other pregnant women at hospital	Yes	Media, hospital	Yes
16	PG001	GDM	29	29	26	Yes	Media	Yes	Media, hospital	No
17	PG009	GDM	26	28	24	Yes	Hospital, husband, friend	Yes	Hospital, husband, friend	No
18	PG010	GDM	21	32	26	No	N/A	Yes	Hospital	No
19	PG011	GDM	25	29	24	Yes	Friend	Yes	Media, friend	Yes
20	PG012	GDM	30	32	24	No	N/A	Yes	Media	Yes

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
21	PG018	GDM	29	28	24	No	N/A	Yes	Media, hospital	No
22	PG021	GDM	28	28	24	No	N/A	No	N/A	Yes
23	PG024	GDM	33	30	26	Yes	Hospital	Yes	Media, hospital	Yes
24	PG025	GDM	23	25	25	Yes	Hospital	Yes	Hospital	Yes
25	PG026	GDM	40	26	25	No	N/A	Yes	Media, hospital, friend	Yes
26	PG028	GDM	26	28	25	Yes	Friend	Yes	Hospital	No
27	PG029	GDM	32	32	24	Yes	Media, hospital	Yes	Media, hospital	Yes
28	PG030	GDM	35	32	29	Yes	Hospital	Yes	Hospital	Yes
29	PG032	GDM	28	32	24	Yes	Media, friend	Yes	Media, hospital, friend	Yes
30	PG033	GDM	32	31	24	Yes	Hospital, other pregnant women at hospital	Yes	Media, hospital, other pregnant women at hospital	Yes
	Mean	**50% GDM**	**28.6**	**29**	**25**	**26.7% No**		**6.6% No**		**83.3% Yes**

4.4.3.2 Interpretation of the Q analysis results

Seven factors were extracted, two were rotated, two were flagged, explaining 39% of the variance and accounting for 22 of the 30 participants. Participant loading of ≥0.456 reached significance at $p < 0.01$, indicating that each loading participant closely exemplifies the factor they load onto.

The factor scores (or factor arrays) for the two flagged factors were the most important aspect of the study result (Table 11). For ease of interpretation it is standard Q-methodological practice to generate a single exemplary Q-sort by merging (according to a procedure of weighted averaging) the Q-sorts of all significantly loading participants: factor array.

The two flagged factors (factor 1 and factor 2) were two significant shared viewpoints on GDM screening among these pregnant women. Each factor pattern (viewpoint) can be exemplified by a best-estimate Q sort (factor scores) in the table. The exemplary Q sort was then interpreted and supplemented by the qualitative comments gathered from participants who had loaded significantly on the factor being interpreted. The interpretative task in Q methodology involves the production of a series of summarising accounts, each of which explicates the viewpoint being expressed by a particular factor (Watts & Stenner, 2005). A clear interpretation of the factor emerges when the various item rankings (scores) and participants comments are effectively combined. The rankings which informed the interpretation at each stage are also included in the interpretation text (as a point of reference for the reader) (Watts & Stenner, 2005).

Factor 1: GDM diagnosis is very important and the OGTT should be conducted for all pregnant women.

Demographic information: Factor 1 has 13 significantly loading participants (participant loading ≥0.456) and it explains 28% of the study variance. It has an eigenvalue of 8.51. There are almost equal number of non-GDM (7) and GDM (6) women in this group. They have an average age of 29.5 years old (Table 13).

Table 13. Basic characteristics of significantly loading participants for factor 1

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
2	P005	Non-GDM	33	26	25	Yes	Friend	Yes	hospital	Yes
4	P007	Non-GDM	24	28	27	Yes	Hospital	Yes	Other pregnant women at hospital	Yes
6	P013	Non-GDM	28	28	24	Yes	Hospital, friend	Yes	Hospital, friend, other pregant women at hospital	Yes
7	P014	Non-GDM	27	27	24	Yes	Media	Yes	Media, hospital	Yes
8	P015	Non-GDM	22	29	25	Yes	Media, other pregnant women at hospital	Yes	Hospital, husband	Yes
9	P016	Non-GDM	32	30	23	Yes	Media, hospital, friend	Yes	Media, hospital	Yes
11	P019	Non-GDM	32	28	24	Yes	Media, hospital, parents	Yes	Media,hospital, parents	Yes
19	PG011	GDM	25	29	24	Yes	Friend	Yes	Media, friend	Yes
25	PG026	GDM	40	26	25	No	N/A	Yes	Media, hospital, friend	Yes

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
26	PG028	GDM	26	28	25	Yes	Friend	Yes	Hospital	No
28	PG030	GDM	35	32	29	Yes	Hospital	Yes	Hospital	Yes
29	PG032	GDM	28	32	24	Yes	Media, friend	Yes	Media, hospital, friend	Yes
30	PG033	GDM	32	31	24	Yes	Hospital, other pregnant women at hospital	Yes	Media, hospital, other pregnant women at hospital	Yes
	Mean	**46.2% GDM**	**29.5**	**29**	**25**	**7.7% No**		**0% No**		**92.3% Yes**

Factor interpretation: This group of women shared a strong viewpoint that testing for GDM is very important and necessary (17: +3) and the OGTT should be conducted for all pregnant women (18: +3). There is neither a burden to undergo the OGTT (23: -3) nor any unhappiness for OGTT (25: -3). OGTT is convenient (1: +1). Having blood drawn three times within 2 hours OGTT is not deemed too much (2: -1), nor is the duration of 2 hours for the OGTT deemed too long (3: -1).

It is very easy to find the scan machine for collecting the GDM diagnosis result (7: +2) and the diagnosis result is understandable (8: -2). Waiting for several hours to collect the OGTT result on that day is better than collecting the OGTT result during the next visit to the hospital (5: -1), since they were eager to know their OGTT results rather than waiting for another 1-2 weeks. Pregnant women did not complain much about the several hours waiting time. It was potentially because in China due to both the massive number of patient visits and concerns over cost savings, the hospitals usually run a large biochemical analyzer to test a batch of hundreds of sample together after these all being collected, it was normal to wait a few hours to collect the result of a test conducted by a large biochemical analyzer (including OGTT).

The whole process of being test for GDM is satisfactory (26: +1), including the support received throughout the OGTT procedure (27: +2). All necessary explanations and advice about OGTT were given by hospital staff (32: +1), and there was no confusion about the OGTT (24: -2). The GDM information received from the doctors/nurses at the hospital was satisfactory (12: -1), including doctor's explanation when giving the GDM diagnosis result sheet to him/her to review (9: +1). Pregnant women are treated with dignity and respect (28: +1). There is no fear of asking the hospital staff relevant advice about OGTT and GDM (31: -2).

The findings suggest that the participants, nevertheless, wished strongly to be given more information about GDM and GDM diagnosis both before (10: +2) and after (11: +2) the OGTT. They also prefer to search GDM information (from internet, TV, magazines, books, etc.) by themselves (18: -2) and to receive GDM information from the hospital bulletin board if available (17: -1). Interestingly, they do not prefer their family (parents, husband/partner) or friends to search for GDM information for them (19: -1).

Factor 2: GDM diagnosis (OGTT) is inconvenient and a burden for pregnant women to some extent.

Demographic information: Factor 2 has 11 significantly loading participants (participant loading ≥0.456) and it explains 11% of the study variance. It has an eigenvalue of 3.36. There are more non-GDM (8) and GDM (3) women in this group. They have an average age of 27.7 years old (Table 14).

Table 14. Basic characteristics of significantly loading participants for factor 2

#	ID	GDM status	Age	Current gestational week	Gestational week for OGTT	Received GDM information before OGTT	Information Source	Received GDM information after OGTT	Information Source	Whether Q study is easy to complete
1	P004	Non-GDM	27	27	23	No	N/A	Yes	Friend	Yes
3	P006	Non-GDM	26	28	24	Yes	Media, friend	Yes	Media, hospital, friend	Yes
5	P008	Non-GDM	25	30	24	Yes	Media, hospital	Yes	hospital	Yes
11	P019	Non-GDM	32	28	24	Yes	Media, hospital, parents	Yes	Media, hospital, parents	Yes
12	P020	Non-GDM	31	30	27	Yes	Friend	Yes	Hospital	Yes
13	P022	Non-GDM	28	28	24	No	N/A	No	N/A	Yes
14	P023	Non-GDM	29	29	24	No	N/A	Yes	Hospital	Yes
15	P027	Non-GDM	29	27	24	Yes	Media, other pregnant women at hospital	Yes	Media, hospital	Yes
21	PG018	GDM	29	28	24	No	N/A	Yes	Media, hospital	No
24	PG025	GDM	23	25	25	Yes	Hospital	Yes	Hospital	Yes
26	PG028	GDM	26	28	25	Yes	Friend	Yes	Hospital	No
	Mean	**27.2% GDM**	**27.7**	**28**	**24**	**36.4% No**		**0.9% No**		**81.8% Yes**

Factor interpretation: This group of women shared a strong viewpoint that OGTT is not convenient at all (1: -3). Having blood drawn three times is too much (2: +3), and the duration of 2 hours for the OGTT is too long (3: +3). However, GDM testing is still important and necessary (21: +1), and OGTT should be conducted for all pregnant women (22: +2). The OGTT does not bring unhappiness (25: -1).

The time between having the OGTT test and collecting the OGTT result is not satisfactory (4: -2). It would be more convenient to collect the diagnosis result at the next hospital visit rather than waiting for several hours to collect it on that day (5: +1). Moreover, compared to collecting the result in person, it is preferable to be informed about the result by telephone when comes out (6: -1). Nonetheless, the scanning machine for result collection is easy to find (7: +2).

The GDM diagnosis results sheet is understandable (8: -1). The GDM information received from the weekly courses for pregnant women at the hospital is not satisfactory (13: -1). Actually, the GDM/OGTT topic was not covered by the weekly courses, though it covered a series of other subjects relevant to pregnancy. The weekly courses were neither a preferred delivery resource for them (15: -1), due to the inflexibility of the timescale of the courses. It is deemed preferable to receive the GDM information from the hospital bulletin board (17: -2).

Throughout the OGTT procedure, the support is satisfactory (27: +1) and pregnant women are respected (28: +1). The doctor who carried out the OGTT listened to pregnant women's concerns and needs (30: +1). Necessary explanations and advice about OGTT are offered (32: +1) and there is no fear at all for asking questions regarding to OGTT and GDM (31: -3). Though satisfied with the GDM information received from the doctors/nurses at the hospital (12: -2) and feels no confusion about

the OGTT (24: -2), it is regarded as strongly desirable to receive more information about GDM and GDM diagnosis both before (10: +2) and after (11: +2) the OGTT.

Comparing factor 1 and factor 2: Viewpoints and participant characteristics

Viewpoints: Both shared viewpoints suggest an equally strong need for more information about GDM and GDM diagnosis both before (10: +2) and after (11: +2) the OGTT.

Factor 1 viewpoint feels OGTT is convenient (1: +1) and not at all a burden (23: -3). However, factor 2 viewpoint feels strongly that OGTT is inconvenient (1: -3) and potentially causes some burden (23: -1). Compared to factor 2 viewpoint, factor 1 believes much more strongly in the importance/ necessity of conducting OGTT (21: +3 for factor 1 and +1 for factor 2), and agree more firmly that OGTT should be undertaken for all pregnant women (22: +3 for factor 1 and +2 for factor 2).

Participant characteristics: When comparing the participants for factor 1 and factor 2, participants for factor 2 had higher proportion of non-GDM women (72.8% versus 53.8%). The factor 2 participants also tended to know less about GDM before OGTT (63.6% versus 92.3% received GDM information). More non-GDM women in factor 2 participants and ignorance of OGTT before undergoing the test might both have contributed to the factor 2 viewpoint that OGTT was inconvenient. Both groups found the Q methodology study easy to complete (81.8% for factor 2 and 92.3% for factor 1).

4.5 DISCUSSION

4.5.1 Statement of principal findings

Two distinctive viewpoints emerged as to whether or not OGTT was convenient. The mainstream viewpoint advanced by the pregnant women was that OGTT was acceptable and not a burden for them (factor 1). However, non-GDM women tended to feel that OGTT was not convenient at all and relatively burdensome (factor 2). More specifically, they felt that having blood drawn three times within 2 hours OGTT was too much, and that the duration of 2 hours was also too long. Although both groups, representing different viewpoints, felt that conducting OGTT was important and necessary for all pregnant women, factor 2 participants ranked this opinion as less important. Overall, non-GDM women tended to feel that OGTT was less important or necessary and reported that OGTT was very inconvenient and relatively burdensome.

Both participant groups ranked highly the statements that they would like to be provided with more information about GDM and OGTT both before and after undertaking the OGTT. Whilst they felt no confusion about the OGTT and agreed that the explanations and advice about OGTT were offered. However, this was only a brief explanation of GDM and their test results when they gave their OGTT result sheet to their doctor for consultation. The pregnant women strongly wished to be provided more information about GDM and OGTT. Regarding the delivery of information, both participant groups preferred to receive information from the hospital bulletin board. Education leaflet were also considered to be an option. Factor 2 participants did not prefer to receive the information from the weekly courses of the “pregnant women’s school” at the hospital, due to the inflexibility of timing.

4.5.2 Strengths and limitations of the study

The representativeness of the study was achieved by involving an equal number of GDM and non-GDM women. The process of conducting the Q methodology study was highly standardised, which avoided potential bias due to different practices. In particular, the study recruited participant who had undertaken their OGTT 2-4 weeks prior to the study. An instruction sheet was provided to each participant to read before they fulfill the FlashQ on the computer. The researcher was in the study room so that the participant could ask any question or help when needed. This also ensured that the FlashQ was accurately completed.

One difficulty of the Q methodology study was the use of exhaustive sources to create a conclusive Q-set of statements. Although this study has developed the Q statements based on various sources including literature reviews, various related websites, five new mothers in the UK, three new mothers in China, and three academic staff at the Warwick Medical School of the University of Warwick, it could still be limited and the Q-set may not be as broad and inclusive of all opinions as ideal. One other weakness of the Q methodology study is that some participants may have found the Q sorting process slightly tiring because of the number of Q statements to sort. The researcher had encouraged their patience for completion before the sorting process started.

4.5.3 Implications for practice

One essential implication of the research results was the need to provide more GDM and OGTT information to pregnant women both before and after OGTT to meet their needs. Both factor 1 and 2 participants indicated a strong need for more information. At the time of the study, only a very brief introduction of OGTT inscribed on the

bottom two lines on the registration card was provided to pregnant women. They either ignored it or felt that the information was insufficient. In line with participant preferences in both groups, the hospital could develop a bulletin board showcasing the information of GDM and OGTT installed on the wall of the waiting area in the Department of Obstetrics and Gynaecology. Another option would be to develop a GDM leaflet to be offered to pregnant women at their first registration or any time before their OGTT. It might not be a priority to include a GDM/OGTT course in the "pregnant women's school" at the hospital, as factor 2 participants in particular pointed out to the inflexibility of the course timing in terms of attendance.

Although both points of view indicated that instructions conveyed by doctor/nurse about OGTT and explanation of GDM were satisfactory, the participants nevertheless wished for further explanation. Based on the findings, doctors are encouraged to give GDM women more information when explaining their OGTT result before referring them to the endocrinology department.

4.5.4 Implications for future research

The non-GDM women tended to rank very highly that the OGTT was not convenient at all, they also felt it was a relative burden for them to some extent. Especially, they felt strongly that having blood drawn three times is too much, and the duration of 2 hours for the OGTT is too long. It is worthwhile considering and exploring ways to reduce the burden of OGTT. As indicated by the current systematic review of the PhD, selective screening might be as effective as universal screening if a risk scoring algorithm was developed and used based on the local population for selecting high risk pregnant women. The burden of non-GDM women could be significantly relieved if a selective screening approach was efficient. However, thus far there was no study assessing the selective screening approach under the IADPSG criteria, thus

the applicability is unknown. It is recommended that such a study be conducted to explore whether a selective screening approach using a risk scoring algorithm will be effective in the Chinese population under the new IADPSG recommendation.

Regarding the Q methodology itself, this is a convenient tool for participants to perform. In the current study, more than 80% participants felt it was easy to complete the Q methodology study. Among the reasons why Q methodology was not convenient, repetitive comments suggested the three boxes for Disagree, Neutral, and Agree were too small when there was a need to drag the Q statements to the Q pyramid. They could not see all the Q statements in the box at a glance but had to scroll the scrolling bar, which was neither convenient or pleasant (Appendix 12). It would be easier for the participants if the three boxes could be enlarged in future FlashQ versions.

4.6 CONCLUSION

Overall, pregnant women agree in different levels that the OGTT for GDM diagnosis is important and needs to be conducted for all pregnant women. Non-GDM women tended to feel strongly that OGTT was inconvenient at all in terms of three blood samples being taken and the 2 hours duration time, and felt that it was a burden relatively speaking to undertake the OGTT. Whist being satisfied with the whole OGTT service and the information received, both GDM and non-GDM women strongly wished to be provided with more information on GDM and OGTT both before and after the OGTT. In practice, hospitals were encouraged to provide more information on GDM and OGTT to pregnant women both before and after OGTT, using hospital bulletin board or by delivering education leaflet. The Q methodology study is a smart and convenient tool to use for investigating perspectives and attitudes.

The findings of the Q methodology study filled the existing gap in knowledge and addressed the research objective of exploring user perspectives on the new IADPSG universal screening approach. This helped to address the overall thesis aim of evaluating and identifying the best screening approach. The Q methodology study results implied the need for future research to explore the possibility of reducing the OGTT burden for non-GDM women, possibly by investigating the effectiveness of a selective screening approach under the IADPSG criteria. In the next chapter, a risk score-based selective screening approach for GDM under the IADPSG will be explored further.

Chapter 5: The effectiveness of a risk score-based selective screening approach for GDM under the IADPSG criteria in China: a case-control study

5.1 BACKGROUND

5.1.1 The IADPSG approach for GDM diagnosis

As stated in Chapter 1, there is no consensus on the best screening and diagnosis approaches for GDM (Tieu *et al.,* 2010; Waugh *et al.,* 2010), whether it should be universal screening or selective screening? Or whether two-step GDM tests or a one-step GDM test should be used? Globally, different countries are implementing different screening approaches for GDM. In 2010, the IADPSG recommended a major change in GDM diagnosis, promoting a one-step 75g OGTT for GDM diagnosis for all pregnant women during 24^{th} to 28^{th} week of gestation with lower threshold values than used previously (Panel, 2010). The new IADPSG approach was expected to substantially increase the GDM incidence, potentially double or triple the incidence (Panel, 2010).

Worldwide controversy continues on whether or not to adopt the IADPSG approach for GDM. China adopted the IADPSG one-step universal approach in July 2011 (Chinese Ministry of Health, 2011), which is the first country applying the IADPSG recommendation. Prior to this time, China implemented a two-step universal approach for GDM, pregnant women underwent a 50g GCT for GDM screening followed by either a 75g or 100g OGTT for confirmation of GDM diagnosis. Diagnosis was made with at least two abnormal values from four measured: fasting glucose ≥5.8 mmol/l, 10.6 mmol/l at 1 hour, 9.2 mmol/l at 2 hours, and 8.1 mmol/l at 3 hours.

Concerns over the costs and cost-effectiveness are among the key reasons that countries hesitate adopting the IADPSG approach. The increased costs are comprised of the increased diagnosis costs by conducting OGTT among all pregnant

women, and the increased healthcare costs by treating the additional women diagnosed as GDM. If, as concluded by all the observational studies, the IADPSG approach is more clinically effective, it will be necessary to spend the additional treatment expenditure on these GDMs. However, there is still a need to explore whether the diagnosis costs of conducting the OGTT could be reduced without comprising the clinical effectiveness, which will improve the cost-effectiveness of the IADPSG approach.

5.1.2 Users perspectives of the IADPSG diagnosis approach

Under the new IADPSG approach, the GDM incidence in China was reported to be from 8% to 15% (Shang & Ma, 2011; Wei & Yang, 2011; Hou *et al.*, 2012; Lu *et al.*, 2012; Jiang *et al.*, 2013). The OGTT diagnostic test has to be performed on 100 pregnant women in order to diagnose 8 to 15 GDM women. Non-GDM women who would have only received a simple 50g GCT test (non-fasting, one hour, one blood sample required) under the two-step approach now have to undertake the more complicated OGTT test (fasting, two hours, three blood samples required). Additional burden has been created on the non-GDM pregnant women by the IADPSG approach.

A Q methodology study of 'Pregnant women's attitudes, views, and experience of the IADPSG approach for GDM in China' was conducted as part of the PhD research (Chapter 4). The result suggested that pregnant women tended to feel a burden to undertake the fasting and 2-hour oral Glucose Tolerance Test (OGTT) for GDM under the IADPSG approach, and they wish more information be given before the test. This result also implied a need to explore the possibility of reducing the OGTTs without comprising the clinical effectiveness, to reduce the burden for pregnant women.

5.1.3 The risk scoring algorithm for selecting high risk pregnant women for GDM diagnosis

Many studies were conducted which compared selective screening versus universal screening approach for GDM. A systematic review of 'The effectiveness and cost-effectiveness of screening for GDM: universal or selective screening?' was conducted as part of the PhD research. The result suggested that although universal screening was recommended for general settings, selective screening could potentially be effective when used self-developed criteria specific for local population for selecting high risk women. In the review, four studies used a self-developed selection criteria based on local population for the selective screening approach (Caliskan *et al*., 2014; Naylor *et al.,* 1997; Pintaudi *et al.,* 2014; Van Leeuwen *et al*., 2010). All achieved satisfactory sensitivity and specificity, and recommended selective screening. Among them, two studies (Naylor *et al.,* 1997; Van Leeuwen *et al*., 2010) used a risk scoring algorithm for selecting women at high risk, which were more specific. These findings implied the possibility that applying a risk scoring algorithm to select high risk women for the IADPSG approach might work.

Whether a selective approach (either by risk factors or by a risk scoring algorithm) is effective or not depends on its sensitivity and specificity. Sensitivity is the proportion of GDM who are diagnosed out by selective approach, while specificity is the proportion of non-GDM who are classified as low risk women and avoid the GDM test. The higher the sensitivity and specificity, the better the selective approach performs. However, the question is, what are our acceptable value of sensitivity and specificity, especially the sensitivity. A sensitivity of 100% by missing no GDM cases is ideal; how about a sensitivity of 90% by missing 10% GDM cases and a sensitivity of 80% by missing 20% of GDM cases? No study capable of answering

this question directly has been found. Some studies investigated the prognosis of missed cases of GDM in selective screening, who were GDM cases with low risk factors. Langer *et al.* (2005) suggested unrecognised GDM in women of normal weight (BMI<25 kg/m^2) was not associated with an increase in the incidence of macrosomia or shoulder dystocia. Cosson *et al.* (2006) found the foetal and maternal prognosis for women with GDM with no risk factor was similar to that of women without GDM. To the contrast, others studies (Moses *et al.,* 1998; Capula *et al.,* 2013) found the pregnancy outcomes of women with GDM from a low-risk group are similar to the outcomes of other women with GDM. There is controversy about whether the prognosis of missed GDM cases is similar to non-GDMs or to other GDM women. A sensitivity of 100% is definitively valuable; however, whether a sensitivity of 90%, 80% or even less is acceptable really depends on each country.

5.1.4 Potential GDM risk factors for establishing the risk scoring algorithm under the IADPSG approach in China

Advanced maternal age, overweight (high BMI), family history of diabetes (among first-degree family member), history of GDM, history of macrosomia, and certain ethnic/racial group including South Asia (especially India, Pakistan or Bangladesh) are commonly recognised risk factors for GDM, as informed by the worldwide guidelines on the risk factors used for GDM selective screening criteria (Association, 1997; Force, 2008; Walker, 2008; NICE 2015) and the risk scoring algorithm studies (Naylor *et al.,* 1997; Phaloprakarn *et al.,* 2009; Van Leeuwen *et al.,* 2010). Other risk factors including gestational weight gain, polycystic ovary syndrome (PCOS), waist circumference, thigh circumference, leg length, mother's birth weight, height, high blood pressure during first trimester, family history of hypertension, irregular menstruation, vulvovaginal candidiasis, triglycerides, total cholesterol, α-Thalassaemia, hemoglobin, maternal smoking, occupation, educational attainment, socio-economic status, carrying Hepatitis B virus are also found to be significantly

associated with developing GDM by previous Chinese studies under the previous two-step approach for GDM diagnosis (Lao & Ho, 2001; Lao *et al.,* 2001; Lao *et al*., 2002; Lao & Ho, 2003; Yang *et al.,* 2002; Jiao, 2003; Yang *et al.,* 2005; Chen *et al*., 2006; Qiang *et al.,* 2006; Lao *et al*., 2007; Ma *et al.*, 2007; Ren *et al.* 2008; Yuan, 2008; Yang *et al.* 2009; Qian, 2012; Du, 2013). In addition, elevated serum ferritin level was found to be a GDM risk factor among the Americans (Chen *et al.,* 2006).

The GDM risk factors are expected to change under the new IADPSG diagnosis approach. With reduced OGTT thresholds, the new IADPSG criteria diagnoses much more women with mild to moderate glucose intolerance (who would not be diagnosed as GDM under the old diagnosis criteria) (Test, 2008).The differences between the GDM and non-GDM women are less distinctive using the IADPSG criteria. As a result, the risk factors might be different or be less than before (odds ratios might decrease and some risk factors might disappear). Only one study investigated the GDM risk factors under the new IADSPG approach in China (Hou *et al.,* 2012). Their study used a retrospective cohort of 1136 pregnant women and explored 14 potential GDM risk factors. The GDM incidence was found be to 10.39%; advanced maternal age, high BMI, family history of diabetes, and polycystic ovary syndrome (PCOS) were identified to be significantly associated with developing GDM (Hou *et al.*, 2012).

5.1.5 The rationale of a risk scoring algorithm study for improving the IADPSG approach in China

The new IADPSG approach for GDM diagnosis was demonstrated to be more clinically effective by all the observational studies (Test, 2008; Lapolla *et al.,* 2011; Benhalima *et al.,* 2013; Shang & Ma, 2011; Wei & Yang, 2011; Lu *et al.,* 2012; Jiang *et al.,* 2013), that the additionally women diagnosed as GDM by the IADPSG

approach had significantly higher incidences of adverse outcomes; diagnosing and treating these women would improve the outcomes. However, concerns exist about the increased costs and the cost-effectiveness of the IADPSG approach (Association, 2010; NIH, 2013). China adopted the IADPSG approach in 2011 (Chinese Ministry of Health, 2011). The intervention and relevant costs for additionally diagnosed GDM women are necessary provided the IADPSG approach is more clinically effective; then exploring the ways to reduce unnecessary diagnosis tests becomes an option.

As implied by the conducted systematic review, the use of a risk scoring algorithm might significantly reduce the screening tests performed without missing much GDM cases (Naylor *et al.*, 1997; Van Leeuwen *et al.*, 2010). However, no study has been conducted to explore the effectiveness of a risk scoring algorithm using the GDM risk factors under the IADPSG approach. The conducted Q methodology study also suggested the need of exploring a selective screening approach to reduce the burden of undergoing OGTT for pregnant women. The current study was conducted to fill the gap and explore whether applying a risk scoring algorithm to select the high risk women for the IADPSG one-step approach would work. A risk scoring algorithm works if it can significantly reduce the OGTT tests performed while still diagnose the same level of GDM cases. By performing less OGTT tests, it might potentially not only reduce the diagnosis costs, but also reduce the burden of non-GDM women for undertaking unnecessary OGTT test.

The specificity of the risk scoring algorithm (i.e., proportion of non-GDM women who exempt from the OGTT tests) were identified at a sensitivity (i.e., proportion of GDM correctly diagnosed) of 100%, 90%, and 80%, respectively. A sensitivity of 100% was ideal, however, the specificity at 90% and 80% sensitivity were also provided for policy making.

5.2. AIM AND OBJECTIVES

The study aimed to investigate the feasibility of using known and novel risk factors for GDM to establish a risk scoring algorithm to identify pregnant women at high risk of developing GDM for the IADPSG approach in China.

The specific objectives were:
(1) to explore the risk factors for GDM under the new IADPSG approach;
(2) to develop and assess a risk scoring algorithm for selecting the high risk pregnant women for IADPSG approach for GDM in China.

5.3. METHODOLOGY

The study included two stages. First, a nested case-control study was conducted to explore risk factors significantly associated with GDM under the IADPSG approach. Second, a risk scoring algorithm (i.e., an equation using adjusted odds ratios of the GDM risk factors to predict the possibility of developing GDM) was developed and assessed for selecting high risk women for the IADPSG approach.

5.3.1 A nested case-control study for investigating the GDM risk factors under the IADPSG approach

5.3.1.1 Nested case-control study design

A nested case-control study was conducted to investigate GDM risk factors and their adjusted odds ratios, in order to formulate the equation of the risk scoring algorithm. A nested case-control study is a type of case-control study conducted within a cohort.

In a nested case-control study, one begins with a defined cohort and identifies cases that have already occurred (retrospective nested case-control study) or as they occur (prospective nested case-control study). Then, for each case, a specified number of controls who have not developed the disease by the time of disease occurrence in the case is selected from the cohort (Ernster, 1994). Similar to a case-control study, the nested case-control study observes from effect to cause. The effect can be a particular condition or disease, while the causes are existing or past attributes or exposures thought to be relevant to the development of the condition or disease under study (Schlesselman & Schneiderman, 1982). The current study used a retrospective nested case-control study design to investigate the GDM risk factors within an existing cohort.

5.3.1.2 Data source

The study used the records of a cohort of pregnant women who delivered in year 2013 at the Chengdu First People's Hospital. The IADPSG one-step universal approach for GDM was implemented by the hospital since the beginning of 2012. The hospital has approximately 3000 birth deliveries per year. Since the GDM incidence was estimated at 8%-15% in China under the new IADPSG approach (Shang & Ma, 2011; Wei & Yang, 2011; Hou *et al.*, 2012; Lu *et al.*, 2012; Jiang *et al.*, 2013), it was estimated that about 300 GDM cases would be identified in year 2013.

5.3.1.3 Potential GDM risk factors to be investigated

Due to the possible dilution or change in GDM risk factors under the IADPSG approach as described in the Introduction section, all risk factors that are potentially

associated with GDM as informed by the previous guidelines and studies were investigated. However, the risk factor of ethnics was not included since the pregnant women would be presumably all Chinese; the thigh circumference and leg length was not investigated as they were not measured at the hospital. A full list of potential GDM risk factors planned for investigation for formulating the risk scores algorithm is shown in Appendix 13. The researcher consulted the Chengdu First People's Hospital and confirmed that data on these variables in Appendix 1 were recorded. The list was further confirmed after the researcher arrived at the hospital and began the process of checking pregnant women's records.

5.3.1.4 Sample size calculation and selection of control group

The number of participants selected to study specific disease-exposure relationships is a fundamental consideration in planning a case-control study (Schlesselman & Schneiderman, 1982). Sample size calculation is based on the anticipated effect size (anticipated differences in risk factors between GDM group and non-GDM group), 80% power and 0.05 significance level (Campbell *et al.,* 1995). Ideally, each risk factor could give an estimate of sample size; the largest sample size should be chosen. In this study, the cohort of about 3000 pregnant women in 2013 was estimated to give about 300 cases. The estimated 300 cases was considered appropriate in sample size, compared to other identified case-control studies on GDM risk factors in China (Lao & Ho, 2003; Chen *et al*., 2006; Yuan, 2008; Qian, 2012; Hou *et al.*, 201; Du, 2013), which used from 21 to 90 cases; and compared to the case-control studies within the two risk scoring systems studies (Naylor *et al.*, 1997; Van Leeuwen *et al.,* 2010), which used 156 and 24 cases, respectively. The other risk scoring algorithm study (Phaloprakarn *et al.,* 2009) investigated the risk factors for a positive 50g GCT screening test, not for a positive OGTT diagnosis test (positive GDM), thus was not considered.

In a nested case-control study, time matching is an essential feature of this design, that one or more controls are matched to cases on date of entry into the cohort, or length of time in the cohort, or a combination of these measures (Wacholder, 1991). The aim of time matching is to avoid the situation that a cohort member who serves as a control at one point in time may later become a case; and that a cohort member may be selected as a control for more than one case (Ernster, 1994). In the current study, since the one-step OGTT took place at the 24th week of gestation at the hospital for diagnosing GDM, all GDM cases and non-GDM controls were automatically matched on the length of time in the cohort. Matching on potential confounding variables is an important consideration of a case-control study design. The primary objective of matching is to eliminate biased comparisons between cases and controls (Schlesselman & Schneiderman, 1982). One advantage of nested case-control study is that cases and controls are automatically matched on factors common to all cohort members (Bailey *et al*., 2005), which are year of admission and hospital of admission in this study. Age, sex, race, weight, occupation, parity, personal or family history of disease are usually served as matching factors. However, in the current study, participants were Chinese females, and the other commonly used matching factors were all potential risk factors for developing GDM under investigation, thus matching on these variables was not considered suitable in the study. Thereby a random selection of non-GDM women with an equal number to the cases (about 300) from the 2013 cohort will be selected as the control group, using randomisation feature in MS Excel.

5.3.1.5 Data collection

The cohort of pregnant women who delivered in 2013 at the hospital was identified from the Electronic Medical Record System (EMRS) at the hospital. This was done by filtering the patients with a hospitalisation date to the Department of Obstetrics

between 1 January 2013 and 31 December 2013 from the EMRS. Their patient IDs were retrieved into a MS Excel sheet. The following records were removed at a later stage: (1) repetitive records of patients who had more than one hospitalisation to the department during the period; (2) patients who were hospitalised but did not deliver at the department (left the hospital before delivery).

GDM cases in 2013 were identified by filtering the International Classification of Diseases (ICD) code of GDM (O24.401) in 2013 in the EMRS. Non-GDM controls were randomly selected in two steps. First, a sub-cohort of pregnant women, with a number that was slightly more than the GDM cases, was randomly selected using the randomisation feature in MS Excel. Second, GDM cases were excluded from the sub-cohort, the remaining ones were served as controls.

The paper version of the patient medical records had all available information of the GDM risk factors. Some of the risk factors were also available in electronic version from the EMRS at the hospital. Data collection was conducted by the researcher in two steps. First, data of risk factors in electronic records were collected. This was done by assessing to the EMRS through an internal computer at the hospital, and browsing the electronic record of each patient. Second, data of the risk factors which were only available in paper records were collected. The patient IDs were used to locate the paper version of their medical records at the Medical Records Room of the hospital. The research was based at the Records Room, took down and browsed the paper record of each patient. Any missing, insufficient, or incorrect data identified from the electronic records were also checked and supplemented by the original paper records.

The list of investigated GDM risk factors and their availability from electronic and paper records is shown in Appendix 13. The whole process was supervised and

facilitated by senior hospital staff in the Department of Gynaecology and Obstetrics, and staff in the Medical Records Room.

5.3.1.6 Data analysis

All the data collected on the MS Excel datasheet were imported into a SPSS datasheet for analysis. IBM SPSS Statistics 21 software was used. The risk factors were independent variables, the GDM diagnosis results was dependant variable. First, univariable analysis was used to identify the risk factors associated with GDM. For each of the GDM and non-GDM group, continuous variables (e.g., age, height, BMI, gestational weight gain, waist circumference) were summaried in terms of means and standard errors (SE); their significances (odds ratio with 95% confidence interval and p-value) were tested using *t*-test. Categorical variables (e.g., family history of diabetes, history of macrosomia, polycystic ovary syndrome, irregular menstruation) were summarised in terms of numbers and proportions; their significances (odds ratio with 96% confidence interval and p-value) were tested using Chi-squared test.

The risk factors which are found to be significantly associate with GDM ($P<0.05$) from the univariable analysis were further analysed by the multivariable analysis (multiple logistic regression analysis) to adjust for interaction and confounding between these risk factors, to identify their independent effect on predicting GDM (i.e., adjusted odds ratio). The multiple logistic regression analysis is used for studies where the dependant variable is categorical variable, thus is appropriate for the study where the GDM diagnosis result (dependant variable) is binary variable (categorical variable). The independent variables (i.e., the risk factors in this study) for multiple logistic regression analysis can be either categorical or continuous variable.

To determine whether continuous variables (i.e., age, height BMI, gestational weight gain, waist circumference) had a linear or categorical (e.g., age≥ 25 years, BMI≥30 kg/m^2) relationship with GDM, the assumption of linearity was evaluated using piecewise polynomials (splines) and visual inspection (Harre*et al.*, 1988). If the relationship was linear, these continuous variables were used directly in the logistical regression analysis. If the relationship was not linear, then thresholds of the variable were identified to approach linearity, transforming the continuous variable into categorical variable (Harre*et al.*, 1988). Finally, all the risk factors data were analysed by the multiple logistic regression to give the adjusted odds ratios (equals to exp(bj) in the logistic model) with 95% confidence intervals and p-values.

5.3.2 Formulation and assessment of the risk scoring algorithm

5.3.2.1 Formulation of the risk scoring algorithm

From the nested case-control study, a list of risk factors which were significantly associated with GDM was identified. The multiple logistic regression result provided their individual effect (adjusted odds ratios) for predicting GDM. The risk scoring algorithm was formulated using the adjusted odds ratios of these risk factors. For example, if six risk factors were found to be significantly associated with GDM from the multiple logistic regression analysis, including age, BMI, gestational weight gain, family history of diabetes, history of macrosomia, and polycystic ovary syndrome (PCOS). The risk score, which is the possibility of developing GDM, equals to exp(A)/[1 + exp(A)]. A=Constant + Log(OR1)×age +Log(OR2)×BMI + Log(OR3)×gestational weight gain + Log(OR4)×family history of diabetes + Log(OR5)×history of macrosomia + Log(OR6) ×PCOS.

5.3.2.2 A ROC curve analysis to assess the risk scoring algorithm

The established risk scoring algorithm gave every pregnant woman a risk score. The higher the risk score was, the bigger chance to develop GDM during pregnancy. To identify an appropriate cut-off score to select high risk pregnant women for IADPSG was essential, which should maximise the sensitivity and specificity of the risk scoring algorithm. The gold standard is usually defined as the "test or criterion" that determines the true presence or absence of the condition (here GDM) in the index-tested population. In this study, the gold standard was a 75g OGTT on all pregnant women (universal approach) in which the high risk population was screened according to a screening criterion. The test being evaluated is usually called the index test. In this study, the index test was screening women to dichotomise into high or low risk of GDM according to the risk scoring algorithm. So a positive index test result was returned if a woman was at high risk, and a negative index test was when the risk scoring algorithm said she was at low risk.

In a standard definition, sensitivity is the proportion of people who have the disease that the test correctly detects. Specificity is the proportion of people who do not have the disease that the test correctly identifies as not having the disease (Bailey *et al*., 2005). In this study, sensitivity was the proportion of GDM who were correctly screened as high risk women by the risk scoring algorithm. Specificity was the proportion non-GDM women who were correctly classified as low risk women and avoided the OGTT under the IADPSG approach.

Different cut-off scores would offer different sensitivity and specificity. A Receiver Operating Characteristic (ROC) curve was used to identify a cut-off score which maximised the sensitivity and specificity of the risk scoring algorithm (i.e., a cut-off score which missed minimal GDM women while exempted a significant proportion

of low risk women from OGTT). A ROC curve is a graphical plot which assesses the overall value of a test if a test is based on an observed variable that lies on a continuous or graded scale (Hanley & McNeil, 1983). The ROC curve is widely used in medical research for assessing the value of a clinical test against the gold standard test, in terms of sensitivity and specificity. In this study, the ROC curve was used to assess the index test of a risk scoring algorithm with the IADPSG approach against the IADPSG universal approach.

IBM SPSS Statistics 21 software was used to draw the ROC curve. Normally, the new test is considered fair if the area under the curve (AUC) is more than 0.7; and the best cut-off score is the most upper-left point. In this study, more weight was given to sensitivity in order to miss as less GDM women as possible. A sensitivity of 100% by missing no GDM was ideal. However, the cut-off score and the corresponding specificity at a sensitivity of 90% and 80% were also acceptable. A specificity higher than 30% was considered effective, which meant the risk scoring algorithm could exempt more than 30% of the non-GDM women from the OGTT test. All the figures at 100%, 90% and 80% sensitivity will be calculated and presented for the study.

5.3.3 Ethical considerations

Pregnant women's records were used to investigate GDM risk factors in this study. However, only their patient IDs were retrieved from the electronic records for identification. Their names, contact information, or any other identifying data were not collected. All women were anonymised and were not identifiable.

All data analysed were collected in a confidential way and anonymity were preserved. Individual women whose records were used in this study were not asked

to provide consent, since there was no requirement for any contact between the women and the researcher; only anonymised data were collected from routinely maintained hospital records.

Ethical consent to conduct this study was sought initially in China, where the Hospital Ethics Approval Committee of the Chengdu First People's Hospital scrutinised and approved the research proposal. A further University of Warwick Biomedical and Scientific Research Ethics Committee (BSREC) approval to conduct this research as part of a Health Sciences PhD programme was sought and granted. The BSREC reference is REGO-2014-705. Please see the full approval letters in Appendix 14.

5.4 RESULTS

5.4.1 Data collection results

5.4.1.1 Records of participants

Ethical approvals from the Chengdu First People's Hospital and the University of Warwick were received in May 2014. Data were collected by the researcher from 1 June to 31 December 2014 at the Chengdu Hospital. A cohort of 2897 pregnant women who delivered at the Department of Obstetrics at the hospital in 2013 was identified from the EMRS at the hospital; their patient IDs were extracted and recorded.

Among the cohort of 2897 women, there were 272 GDM cases. The GDM incidence was 9.4%. A sub-cohort of 310 women was randomly selected using the

randomisation feature in MS Excel. After excluding the GDM women in the sub-cohort, 278 non-GDM women formed the control group.

5.4.1.2 Records of potential GDM risk factors availability of records

Electronic and paper records of these 550 women were retrieved according to the list of 26 potential GDM risk factors under investigation. Details of historical risk factors were supplied by the pregnant women and recorded by the clinicians at the Department. All the current risk factors were tested or measured either by the Chengdu Hospital or occasionally by another hospital if the women were transferred to the Chengdu Hospital from another hospital during her pregnancy.

The availability of the records is shown in Appendix 13. Gestational weight gain before 24th gestational week, PCOS, waist circumference at first antenatal visit, and blood pressure during first trimester were only available through paper records. For risk factors which were available through electronic records, some had missing data and were supplemented by paper records. These involved family history of diabetes, history of macrosomia, family history of hypertension, vulvovaginal candidiasis, hemoglobin during first trimester, serum ferritin level during the first trimester, and HBV. Personal history of GDM, mother's birth weight, triglycerides, total cholesterol, α-thalassaemia, education level, and family income were not available from the records. Maternal smoking was recorded as 'No' for all the 550 women. After the first round of data collection, missing data and outliers were checked and were supplemented or corrected by browsing the relevant records for a second time.

5.4.2 GDM incidence and basic characteristics of participants

272 GDM cases were identified from the cohort of 2897 pregnant women of in 2013. GDM incidence was 9.4%. The basic characteristics of participants are shown in Table 15. The mean age of GDM women and non-GDM women was 30 and 28 years old, respectively. The average height of GDM women and non-GDM women was 158.5 cm and 159.8 cm. The GDM group had higher BMI (22.5 versus 21.2) and larger waist circumference (85.1cm versus 82.6cm) on average as compared to non-GDM women. The unemployment rate was equally low in the two groups, which was 3.4% for GDM women and 3.3% for non-GDM women. About 34.7% of the GDM women and 20.5% of non-GDM women had previous deliveries. Also, 8.1% of the GDM women and 9.4% of the non-GDM women had a previous adverse history of pregnancy. Within the GDM group, 11.4% women had a family history of diabetes and 21.7% women had a family history of hypertension, whereas within the non-GDM group, it was 4.0% and 23.4% respectively.

Table 15. Basic Characteristics of participants

	GDM (n=272)	**Non-GDM (n=278)**
Age	29.8 ±5.1	27.6 ±3.9
Height (cm)	158.5 ±5.2	159.8 ±4.4
BMI (kg/m2)	22.5 ±3.4	21.2 ±3.0
Waist circumference (cm)	85.1 ±8.6	82.6 ±7.9
Occupation (unemployed)	3.4%	3.3%
Previous deliveries	34.7%	20.5%
Adverse pregnancy history	8.1%	9.4%
Family history of diabetes	11.4%	4.0%
Family history of hypertension	21.7%	23.4%

5.4.3 Formulation of the risk scoring algorithm

5.4.3.1 GDM risk factors identified from univariable logistic regression

Univariable logistic regression was performed to test the association between each risk factor and GDM, using the binary logistic regression feature in IBM SPSS Statistics 21. Table 16 shows the result of the univariable logistic regression. Factors with a P-value of less than 0.05 were potential risk factors associated with GDM, and have been indicated in bold letters in Table 16. Age, height, BMI, previous deliveries, family history of diabetes, systolic and diastolic blood pressure during first trimester, and waist circumference were identified as potential GDM risk factors.

Table 16. Univariable logistic regression result for identifying potential GDM risk factors

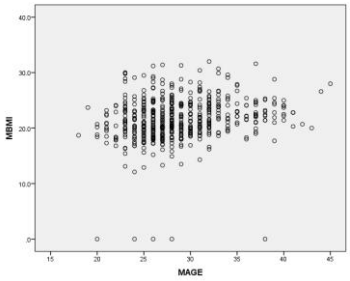
	GDM (n=272)	Non-GDM (n=278)	P-value	Odds Ratio
Age	29.8 ±5.1	27.6 ±3.9	**0.000**	1.116
Height (cm)	158.5 ±5.2	159.8 ±4.4	**0.002**	0.945
BMI (kg/m2)	22.5 ±3.4	21.2 ±3.0	**0.000**	1.120
Waist circumference (cm)	85.1 ±8.6	82.6 ±7.9	**0.000**	1.040
Gestational weight gain at 24th week (kg)	6.5 ±4.5	6.3 ±5.2	0.510	1.012
Occupation (unemployed)	3.4%	3.3%	0.942	1.035
Previous deliveries	34.7%	20.5%	**0.000**	2.243
Adverse pregnancy history	8.1%	9.4%	0.600	0.853
History of macrosomia	4.2%	3.7%	0.736	1.162
Irregular menstruation	3.7%	5.8%	0.156	0.573
Vulvovaginal candidiasis (VC)	10.4%	9.3%	0.684	1.130
Hepatitis B virus carrier (HBV)	8.9%	6.2%	0.233	1.480
Family history of diabetes	11.4%	4.0%	**0.002**	3.122
Family history of hypertension	21.7%	23.4%	0.635	0.908
Polycystic ovary syndrome (PCOS)	0.7%	0.0%	0.999	1657357548
Hemoglobin (>130g/L)	21.7%	20.8%	0.806	1.071
Serum ferritin level (≥upper quartile of 123.5ng/ml)	26.2%	24.1%	0.916	1.038
Systolic blood pressure (≥120mmHg)	29.7%	19.4%	**0.017**	1.900
Diastolic blood pressure (≥80mmHg)	33.8%	14.9%	**0.002**	3.429

5.4.3.2 The risk scoring algorithm established from the multiple logistic regression

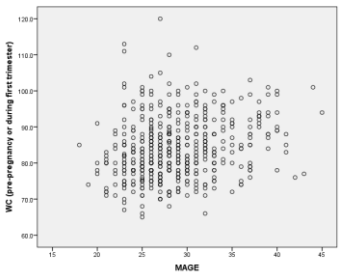
Before performing multiple logistic regression, the correlations between each of the potential risk factors were tested using SPSS. Co-linearity between two continuous variables was tested by scatter plot. Correlation between a categorical and a continuous variable was tested by means of a chi-square test. Correlation between two categorical variables was not applicable in this study; if applicable, it would be tested by box plot. The overview of the correlation result is illustrated in Table 17. The detailed correlation results are shown in Figure 14, 15, 16.

Table 17. Overview of co-linearity between each of the potential risk factors

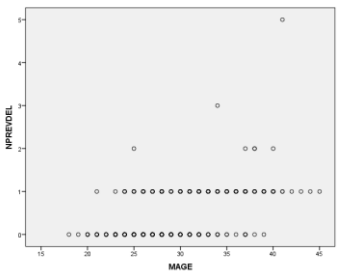
Risk factor	Co-linearity/correlation
1. Age and height	Not applicable
2. Age and BMI	No
3. Age and waist circumference	No
4. Age and previous deliveries	No
5. Age and family history of diabetes	Not applicable
6. Age and systolic blood pressure	No
7. Age and diastolic blood pressure	No
8. Height and BMI	No
9. Height and waist circumference	No
10. Height and previous deliveries	Not applicable
11. Height and family history of diabetes	No
12. Height and systolic blood pressure	No
13. Height and diastolic blood pressure	No
14. BMI and waist circumference	**Yes**
15. BMI and previous deliveries	No
16. BMI and family history of diabetes	No
17. BMI and systolic blood pressure	No
18. BMI and diastolic blood pressure	No
19. Waist circumference and previous deliveries	No
20. Waist circumference and family history of diabetes	No
21. Waist circumference and systolic blood pressure	No
22. Waist circumference and diastolic blood pressure	No
23. Previous deliveries and family history of diabetes	No
24. Previous deliveries and systolic blood pressure	Not applicable
25. Previous deliveries and diastolic blood pressure	Not applicable
26. Family history of diabetes and systolic blood pressure	Not applicable
27. Family history of diabetes and diastolic blood pressure	Not applicable
28. Systolic blood pressure and diastolic blood pressure	**Yes**



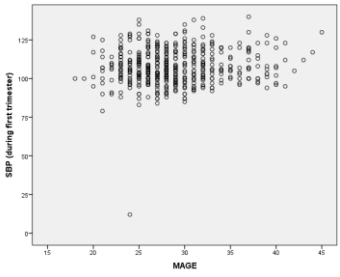
2. Age and BMI



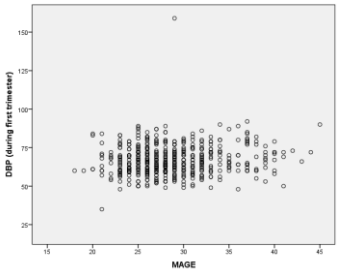
3. Age and waist circumference



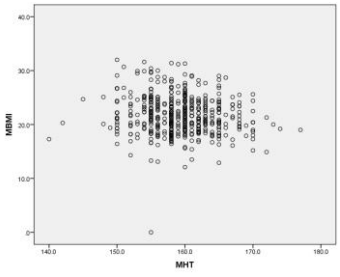
4. Age and previous deliveries



6. Age and systolic blood pressure

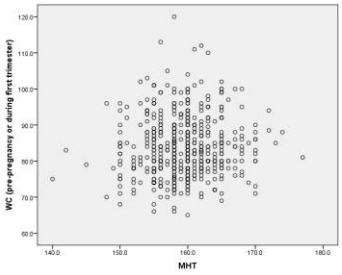


7. Age and diastolic blood pressure

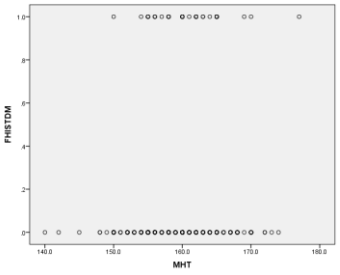


8. Height and BMI

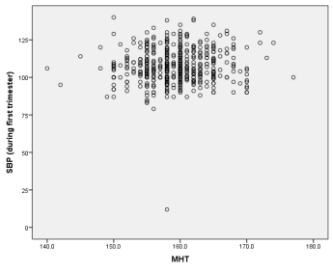
Figure 14. Scatter plot result for the co-linearity/correlation between two continuous variables



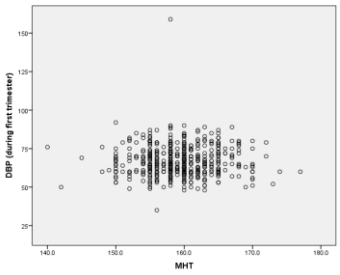
9.Height and waist circumference



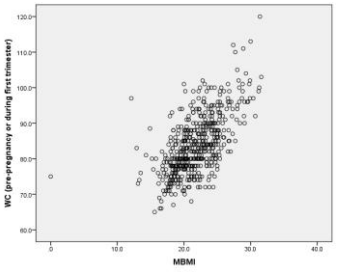
11.Height and family history of diabetes



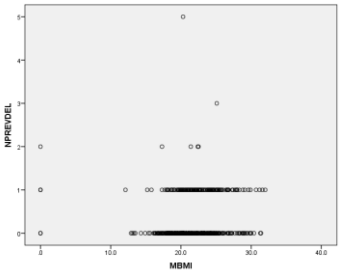
12. Height and systolic blood pressure



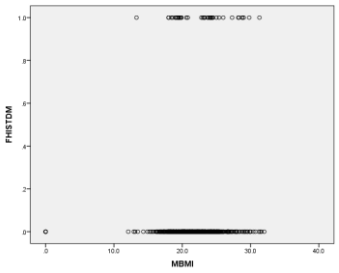
13. Height and diastolic blood pressure



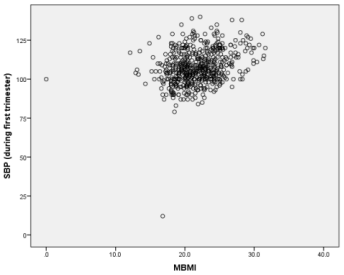
14. BMI and waist circumference



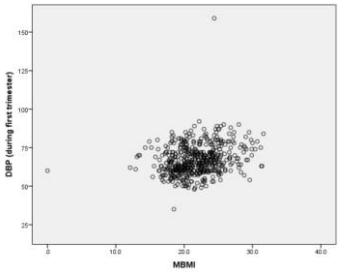
15. BMI and previous deliveries



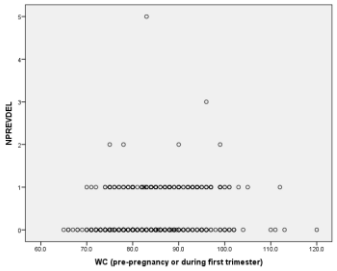
16. BMI and family history of diabetes



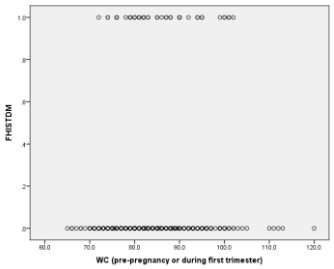
17. BMI and systolic blood pressure



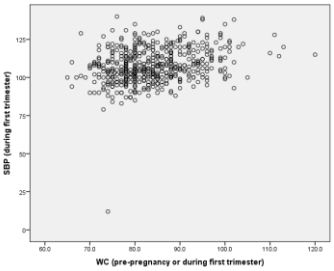
18. BMI and diastolic blood pressure



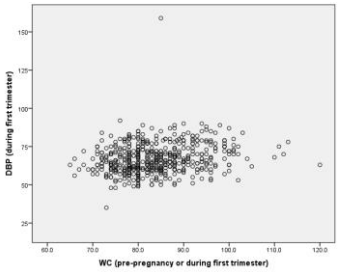
19. Waist circumference and previous deliveries



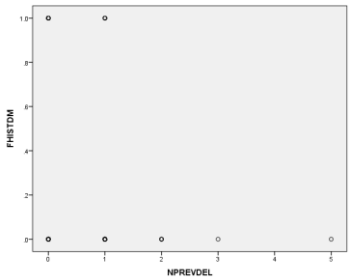
20. Waist circumference and family history of diabetes



21. Waist circumference and systolic blood pressure



22. Waist circumference and diastolic blood pressure



23. Previous deliveries and family history of diabetes

Figure 14 (continued). Scatter plot result for the co-linearity/correlation between two continuous variables

Correlations

		MBMI	WC (pre-pregnancy or during first trimester)
MBMI	Pearson Correlation	1	.625**
	Sig. (2-tailed)		.000
	N	550	530
WC (pre-pregnancy or during first trimester)	Pearson Correlation	.625**	1
	Sig. (2-tailed)	.000	
	N	530	530

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 15. Correlation between BMI and waist circumference as identified from the scatter plot

SBP(if>120) * DBP (if>80) Crosstabulation

Count

		DBP (if>80)		Total
		0	1	
SBP(if>120)	0	448	15	463
	1	46	22	68
Total		494	37	531

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	77.524[a]	1	.000		
Continuity Correction[b]	73.098	1	.000		
Likelihood Ratio	50.473	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	77.378	1	.000		
N of Valid Cases	531				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.74.

b. Computed only for a 2x2 table

Figure 16. Chi-square test result for the correlation between two categorical variables (between systolic blood pressure and diastolic blood pressure)

As illustrated in Table 17 as an overview, BMI and waist circumference were correlated, as well as systolic and diastolic blood pressure. For correlation between BMI and waist circumference (Figure 15), the p-value for the Pearson correlation test was less than 0.01, which suggested that they were significantly associated with each other. The correlation between systolic and diastolic blood pressure was shown in Figure 16. Again, the p-value for the Pearson correlation test was less than 0.01, indicating that systolic and diastolic blood pressure was correlated with each other.

Although the correlation analysis showed two of the eight GDM risk factors were not independent risk factors for GDM, to maximise the accuracy of the predication, all the eight risk factors needed to be added to the multiple logistic regression model to establish the risk scoring algorithm. To predict the possibility of developing GDM, it is not necessary to know the individual independent effect of each risk factor, but to have the overall effect of these risk factors for prediction.

Several performances were also made to compare the ROC curves using all eight risk factors and all non-correlated factors (six risk factors). The ROC curve using all eight risk factors achieved much better sensitivity, specificity, and area under the curve. This also supported the rationale of entering all risk factors into the multiple logistic regression model for prediction.

Table 18. Multiple logistic regression result for formulating the risk scoring algorithm

	B	P-value	Odds ratio
Age	**0.069**	0.007	1.072
Height	**-0.058**	0.006	0.944
BMI	**0.06**	0.11	1.061
Waist circumference (WC)	**0.009**	0.557	1.009
Previous deliveries (PD)	**0.38**	0.117	1.462
Family history of diabetes (FHD)	**1.163**	0.003	3.198
Systolic blood pressure (SBP) (≥120mmHg)	**0.308**	0.326	1.361
Diastolic blood pressure (DBP) (≥80mmHg)	**0.952**	0.031	2.592
Constant	**4.87**	0.152	130.376

The risk score, which is the probability of developing GDM is formulated as： Risk score =probability =exp (A)/ [1 + exp (A)]; while A =4.87 + 0.069*Age - 0.058*Height + 0.06*BMI + 0.009*WC + 0.38*PD + 1.163*FHD + 0.308*SBP + 0.958*DBP (see Table 18 for abbreviations). Although BMI, waist circumference, previous deliveries, and systolic blood pressure became insignificant in the multiple logistic regression, they were still predicators for developing GDM. To make the risk scoring algorithm simpler, the last decimal in the formula was rounded up and it was tested that there was no difference. The final risk score was calculated as: Risk score =probability =exp (A)/ [1 + exp (A)]; while A =4.87 + 0.07*Age - 0.06*Height + 0.06*BMI + 0.009*WC + 0.38*PD + 1.16*FHD + 0.31*SBP + 0.96*DBP (see Table 18 for abbreviations).

Linearity analysis of the continuous variables (age, height, BMI, waist circumference) was conducted using visual inspection (Harre*et al.*, 1988). The relationships between these continuous variables and GDM were linear, which suggested that these continuous variables could be used directly in logistical regression analysis. For the same reason, the continuous variables (age, height, BMI, waist circumference) were not transformed into categorical variables using cut-off thresholds, which could significantly reduce the accuracy and effectiveness of the risk scoring algorithm.

5.4.4 Assessment of the risk scoring algorithm

Using the established risk scoring algorithm, each woman's risk score (0~1) was calculated based on her risk profile. A ROC curve as drawn using SPSS (Figure 17). The vertical axis is sensitivity, which is the proportion of GDM women who are correctly being categorised as high-risk by the risk scoring algorithm and are diagnosed. The lateral axis is 1-specificity, that specificity is the proportion of non-GDM women who are correctly categorised as low-risk by the risk scoring algorithm

and avoid OGTT. The curve is the risk score, by which different cut-off scores give different combinations of sensitivity and specificity.

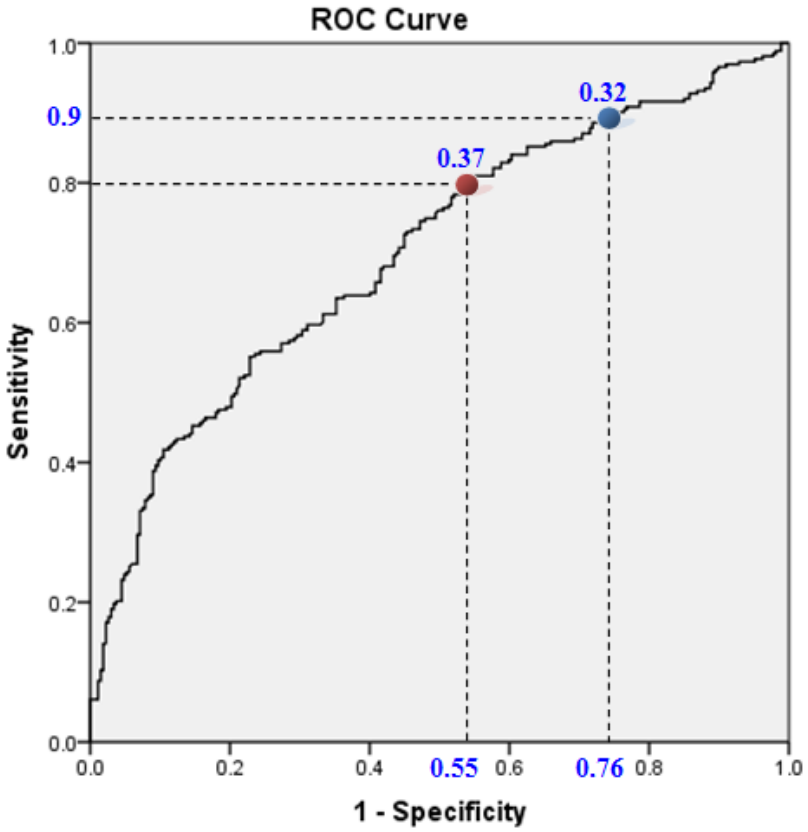


Figure 17. ROC curve with cut-off scores for assessing the risk scoring algorithm

The Area under curve (AUC) is 0.7, which is fair for an index test (the risk scoring algorithm). A full list of cut-off scores and corresponding sensitivities and specificities of the ROC curve was illustrated in Appendix 15. Table 19 shows the two cut-off scores identified at 80% and 90% sensitivity of the risk scoring algorithm. A cut-off score of 0.37 (the red point in Figure 17) was identified at a sensitivity of 80%, and the corresponding specificity is 45%. This means using a cut-off score of 0.37 for selecting high risk women can exempt nearly half (45%) of non-GDM

women from the OGTT, and still diagnose 80% of GDM. A cut-off score of 0.32 (the green point in Figure 17) was identified at a sensitivity of 90%, with the corresponding specificity being 24%. This indicates that using a cut-off score of 0.32 could miss less GDM women (only 10%), but could only exempt 24% of the non-GDM women from test. A more detailed 2x2 table showing the figures behind the two cut-off scores was illustrated in Appendix 16.

Table 19. Cut-off scores for 80% and 90% sensitivity of the risk scoring algorithm

Risk score	Specificity	Sensitivity
0.37	**45%**	**80%**
0.32	24%	90%

There are no definitive cut-off points or standards for determining what level of sensitivity and specificity that can indicate effective and suitable screening, especially in view of the trade-off between sensitivity and specificity. However, a number of factors can be considered to determine the levels of sensitivity or specificity that are acceptable. First of all, the context in which screening has to take place plays an important role in decision making. For GDM screening, as in this study, we are more interested in sensitivity with the aim of missing as few GDM women as possible than specificity due to the potential for adverse events in GDM women and newborns. Second, the suitability of a screening test also takes into consideration the simplicity and costs. The risk scoring algorithm as an index test in this context is quick, cheap and easy, enabling doctors to calculate the risk scores of pregnant women based on their existing medical records. On the other hand, OGTT as the actual test, is relatively complicated in terms of the fasting state required and the need for two hourly tests as well as three blood samples. Thus, a high specificity to avoid unnecessary OGTT would be important and favorable. Third, whether or not a type of screening should be adopted also depends on the size and the condition of the population. Pregnant women make up a large group of the population, and they

are in a vulnerable condition. Therefore, high specificity to exempt mass non-GDM women from the screening would be advantageous. Based on the reasons outlined above, a sensitivity of 80% was deemed to be acceptable with a specificity of 45%. A selective IADPSG approach using a risk scoring algorithm based on local population profile was potentially effective in China.

5.5 DISCUSSION

5.5.1 Statement of principal findings

The incidence of GDM in China was 9.4% under the new IADPSG screening criteria. A selective screening approach to screening for GDM was developed using a risk score. The strategy hinged on a clinical scoring algorithm that grouped pregnant women according to their risk of developing GDM. Its efficiency was achieved by not screening pregnant women in the low risk group, thus significantly reducing the number of unnecessary OGTT. The risk scoring algorithm was able to spare nearly half (45%) of the non-GDM women the need to undergo a screening blood test (OGTT), which composed 41% of the whole pregnant population. Use of the risk score-based selective screening approach could save numerous monetary costs and distress for individual pregnant women in China. However, on the other hand, 20% of GDM women (that is 2% of the whole pregnant population), would be missed if a selective screening approach is used, who might develop adverse clinical outcomes.

5.5.2 Strengths and limitations of the study

The nested case-control study design has several strengths. It allows the study of multiple potential causes of a disease (e.g., multiple risk factors for developing GDM); it is useful for the study of rare cases (e.g., GDM with an incidence of 9.4% in China under the IADPSG approach). The retrospective nested case-control study uses existing data and avoids effort devoted to the follow-up of individuals. It is free from selection bias (unable to refuse for participation) or differential recall bias (all data were examined and recorded rather than self-reported). Compared to case-control study, the advantage of nested case-control study is that the cases and controls are from the same cohort, thereby automatically matched on factors common to all cohort members and reduces bias (Bailey *et al.*, 2005).

The study was advantageous in exploring a full range of potential GDM risk factors (more than 20 risk factors) which were more extensive than the previously conducted two risk score studies (Naylor *et al*., 1997; Van Leeuwen *et al.*, 2010). Naylor *et al*. (1997) explored six risk factors involving age, race, BMI, parity, family history of diabetes, adverse obstetrical history. Van Leeuwen *et al.* (2010) explored eight risk factors which were age, BMI, ethnicity, family history of diabetes, smoking, previous miscarriage, history of GDM, history of perinatal death.

The disadvantage might be that the case-control study design relies on existing patient records, which sometimes can be insufficient, inaccurate, or impossible for validation. In the present study, all participants answered 'No' for smoking during pregnancy, and one might question the reliability o this. Another limitation is that there were no data on the GDM history of patients which was thought to be a significant risk factor for developing GDM. The lack of data was due to China's one-child policy from 1979 to 2013, whereby couples in China could only have one child. Ethnic minorities are exceptions, and also in some rural areas, families were allowed to have two children if the first was a girl. Since 2013, the policy changed, in that

couples where at least one was a single child could have two children. Since 2015, China has completely ended the one-child policy and couples are now allowed to have two children. In this study, the collected data showed that 20.5% non-GDM women and 34.7% GDM women had previous deliveries. A certain proportion of these women could have a history of GDM which were not recorded. The risk scoring algorithm could be more accurate and effective if the GDM history data of these women were recorded.

In addition, the setting of the retrospective case-control study was potentially different when applying the risk score for selective screening in a prospective setting. Furthermore, the risk score algorithm was established and assessed using the retrospective cohort of 2897 pregnant women who delivered at the hospital in 2013 (derivation cohort). The effectiveness of the risk scoring algorithm might need further validation especially using a prospective cohort (validation cohort) from another district in China.

5.5.3 Implication for practice and further research

By way of implications for practice, a selective screening approach using a risk score might be considered an alternative option to the IADPSG universal screening under implementation in China. The decision to choose the selective approach is subject to consideration of the tradeoff between saved costs (monetary costs, time cost, distress for individuals) for nearly half of the non-GDM women (about 41% of pregnant women) and the potentially adverse outcomes of overlooking 20% GDM women (about 2% of pregnant women). Although not all missing GDM women will develop adverse clinical events, if this happens, the outcome will be unequivocally adverse for any woman who does develop GDM.

Clinicians will not find it difficult to use the scoring algorithm. All eight risk factors can be found easily on the patient records. If applied in real practice, a simple computer or web-based calculation tool similar to QRISK2 tool for cardiovascular risk calculation (https://www.qrisk.org/2016/) might need to be developed to enhance the ease of application.

In terms of the research implications, a further validation cohort is needed to validate the effectiveness of this risk scoring algorithm. More specifically, a prospective cohort from other areas of China would be preferable. It is recommended that the effectiveness of such a risk scoring algorithm should be assessed by other countries who currently using a selective screening approach based on standard risk factor guidelines as well as countries, which while implementing universal screening are considering the possibility of a selective screening approach. For countries that have already adopted the IADPSG universal screening approach, an effective risk scoring algorithm for selective screening might potentially save considerable costs and distress for pregnant women.

5.6 CONCLUSION

A risk scoring algorithm for selective screening based on the local Chinese population was established and assessed. It was able to exempt 45% of non-GDM pregnant women from undergoing OGTT, but it overlooked 20% of the GDM women. Whether or not to implement a universal or selective screening approach depends on consideration of the tradeoff between burdening exemptible non-GDM women (41% of whole pregnant population) and the possible adverse clinical events of the overlooked GDM women (2% of whole pregnant population). A validation

study using another cohort will be needed to further affirm the effectiveness of the risk scoring algorithm.

The findings of the study addressed the research objective of establishing and evaluating the effectiveness of a risk score-based selective screening approach under the IADPSG criteria. The results suggested that it could be considered as an alternative screening approach with the added benefit of significant monetary savings and burden exemptions. This was due to the fact that it would spare nearly half of the pregnant women from undergoing unnecessary OGTT, but the cost of overlooking 20% GDM women (2% pregnant women) would still exist. The findings supported the overall thesis aim with reference to exploring the optimal screening approach for GDM.

Chapter 6: Overall Discussion

6.1 AN OVERVIEW OF MAIN FINDINGS

GDM prevalence has increased with the increasing global epidemic of obesity and type 2 diabetes, as well as other factors, including increased maternal age. The GDM prevalence varies significantly among countries because of different population characteristics and risks, screening approaches, screening tests, and adopted criteria.

The evaluation and development of an effective and cost-effective screening approach for GDM is invaluable for pregnant women, their offspring, and the whole society. This research evaluated the current evidence on universal versus selective screening, investigated pregnant women's perspective for GDM screening, and explored the effectiveness of a selective screening approach for GDM based on risk scores in Chinese population under the new IADPSG criteria. Several important inconsistencies in evidence and research gaps have been answered and fulfilled, thereby providing several implications of the current practice in GDM screening. However, certain challenges are still encountered. Future studies are recommended on risk score-based selective screening and qualitative perspective research under other local screening settings other than China to facilitate more sophisticated and well-informed decision making in a particular country.

6.1.1 Universal versus selective screening for GDM

The systematic review included 33 studies, involving 28 effectiveness studies, four cost studies and one cost-effectiveness study. A trade-off was observed between sensitivity and specificity for the effectiveness of selective screening. Only seven of the 28 effectiveness studies recommended selective screening for pregnant women (Caliskan *et al*., 2014; Helton *et al*., 1997; Jensen *et al*., 2013; Naylor *et al.,* 1997; Sacks *et al*., 1987; Van Leeuwen *et al.*, 2010; Pintaudi *et al*., 2014). Compared with

studies which recommended universal screening, these seven studies had been conducted in areas of relatively low GDM prevalence. The seven studies included all the four studies that had developed their own selection criteria for identifying high risk women based on local population risk profiles rather than on standard guidelines (Caliskan *et al*., 2014; Naylor *et al.*, 1997; Van Leeuwen *et al*., 2010; Pintaudi *et al.*, 2014). Two of the four studies used a risk scoring algorithm (Naylor *et al.*, 1997; Van Leeuwen *et al*., 2010). This suggested that selective screening with self-developed selection criteria for high risk women could potentially be effective.

Limited evidence exists on the cost-effectiveness of selective screening compared to universal screening. The cost-effectiveness study (Poncet *et al*., 2002) and three of the four cost studies (Coustan *et al.*, 1989; Larijani *et al.*, 2004; Shamsuddin *et al*., 2001) showed that universal screening was slightly more expensive and less cost-effective in comparison with selective screening.

6.1.2 Pregnant women's perspectives of GDM screening

China is currently implementing the new IADPSG one-step universal screening approach for GDM. Q methodology was used to investigate the perspectives of Chinese women on GDM screening under this approach. Thirty pregnant women (15 GDM and 15 non-GDM women) were involved in ranking the pre-developed 32 Q statements. Two distinctive points of view shared by the respondents emerged from the Q analysis results. In general, the pregnant women agreed that GDM screening was important and necessary to be conducted for all pregnant women. However, non-GDM women felt strongly that GDM screening test (OGTT) was inconvenient in terms of the three blood samples needed and the duration of two hours, thereby experiencing a relative burden in undergoing the OGTT. Both GDM and non-GDM

women keenly desired to be provided with more information on GDM and OGTT both before and after undergoing the GDM screening.

Although it is useful to learn that universal screening is generally accepted by pregnant women, the result also illustrates the importance of exploring an effective selective screening approach that might exempt non-GDM pregnant women without affecting GDM women. In addition, since this has been identified as a strong need, more information on GDM and OGTT should be offered to pregnant women.

6.1.3 The effectiveness of risk score-based selective screening under IADPSG

A risk score-based selective screening approach for GDM was developed and assessed under new IADPSG criteria in China. A risk scoring algorithm for identifying high risk pregnant women was established, using existing and novel risk factors identified from the nested case-control study. By using a cut-off score of 0.37, the risk score-based selective screening approach yielded a sensitivity of 80% and a specificity of 45%. . Using this approach, on one hand, 45% of the non-GDM, which is 41% of the whole pregnant population were correctly identified as low-risk women and thereby spared from the unnecessary OGTT screening test. While this indicates potential for significant monetary savings for the country and prevention of unnecessary burden for pregnant women, this approach carries possible risks. The advantages were available at the cost of missing 20% of the GDM women, which percentage represented 2% of the pregnant population, thereby highlighting concerns when making a decision as to the adoption of the optimal screening approach.

6.2. COMPARISON WITH PREVIOUS STUDIES

6.2.1 Universal versus selective screening for GDM

Two systematic reviews (Hieronimus & Le Meaux, 2010; Tieu *et al.*, 2010) and two health technology assessment (HTA) reports (Scott *et al.*, 2002; Waugh *et al*., 2010) have been conducted previously, involving a comparison of universal versus selective screening. They were limited due to either the databases searched (Hieronimus & Le Meaux, 2010), the types of study included (Tieu *et al.*, 2010), or the focus and number of the effectiveness studies included (Scott *et al.*, 2002; Waugh *et al*., 2010).

Hieronimus and Le Meaux (2010) used efficacy outcome measures of sensitivity and specificity for effectiveness studies. However, they searched only two databases of Medline and Cochrane database from 1990 to 2010, and did not assess the quality of the included studies. A total of 14 effectiveness studies were included in the review, which concluded that the benefit of GDM screening and treatment were only proven in women presenting GDM risk factors, while the relevance of screening for women with no risk factors remained controversial. Tieu *et al.* (2010) evaluated the screening, diagnosis and treatment of GDM, and included only randomised and quasi-randomised trials. Only one quasi-experimental study (Griffin *et al.*, 2000) was identified comparing the clinical outcome measures of the two screening approaches, which in itself reflects bias and limitations. Tieu *et al.* (2010) did not make a definite conclusion about which screening approach should be recommended. Scott *et al.*, (2002) and Waugh *et al*. (2010) evaluated the treatment and screening for GDM. Each of these reviews involved a short sub-section discussing risk factor screening compared to universal screening. Only nine effectiveness studies were described in the review by Scott *et al.*, indicating that selective screening on the basis of risk factors would miss about half of the women with GDM. Waugh *et al*. (2010)

included an additional study by Cosson *et al*. (2006) with regards to the effectiveness of selective screening versus universal screening to the review by Scott *et al*.

The current systematic review extended and updated the previous reviews by using broader types of study designs, including the recent studies published after 2010 as well as fully analysing the efficacy outcome measures of sensitivity and specificity of selective screening in comparison with universal screening. This is the most comprehensive and up-to-date systematic review on universal versus selective screening up till this point, and a number of informative conclusions as to the two approaches have been drawn in comparison with those of the four preceding reviews.

6.2.2 Pregnant women's perspectives of GDM screening

In line with the participant views expressed in the single study done by Griffiths *et al.* (1993) in Australia which explored the attitudes of pregnant women toward universal one-step GDM screening with modified OGTT, the participants of this current study also generally felt it was important and necessary to conduct a screening test for all pregnant women. However, unlike positive attitudes towards the convenience of the screening method used in Australia, non-GDM pregnant women in China tended to feel that the OGTT test was not convenient at all, and that it was potentially burdensome. This may be attributable to the difference in the GDM screening tests administered in the Australian study and those administered in the current study in China. In the Australian study, a 75g OGTT involved one blood sample taken in a fasting state either at home or in a collection center, followed by a further blood sample taken after 2 hours. Whereas in China, a 75g OGTT involved one blood sample taken in a fasting state at the hospital, followed by two blood samples taken after 1 hour and 2 hours. The 75g OGTT was more complicated in the current study in China. Apart from the differences in OGTT itself, there might be

other contributors to their attitudes including cultural issues and knowledge about OGTT before undergoing the test. The Q methodology study showed that 63.6% of factor 2 participants and 92.3% of factor 1 participants received GDM information before undergoing OGTT. Health education about GDM and OGTT before the test would make the OGTT process more acceptable.

6.2.3 The effectiveness of risk score-based selective screening under IADPSG

Four previous studies developed and used their own developed selection criteria for identifying high risk women for GDM screening (Caliskan *et al*., 2014; Naylor *et al*., 1997; Van Leeuwen *et al*., 2010; Pintaudi *et al.,* 2014). Two of these used a risk scoring algorithm and were conducted in Canada (Naylor *et al.,* 1997) and the Netherlands (Van Leeuwen *et al*., 2010) respectively. Naylor *et al*. (1997) found that risk score-based selective screening achieved a sensitive of 90.6% and specificity of 34.7%, while Van Leeuwen *et al.* (2010)'s risk scoring algorithm reached a sensitivity of 75.0% and a specificity of 57.0%. Both studies recommended selective screening over universal screening as a conclusion.

The findings of the current risk scoring study in China were consistent with and supported the two previous risk scoring studies. It found that the risk scoring algorithm yielded a sensitive of 80% and a specificity of 45%. It may be noticed that there were differences in setting among the three studies. The Canadian study used two-step GDM tests and had a GDM prevalence of 2.1% in 1997 (Naylor *et al.,* 1997). The Netherlands study used two-step GDM tests and had a GDM prevalence of 4.6% in 2010 (Van Leeuwen *et al*., 2010). For the current study in China, the one-step GDM test using IADPSG criteria was used and the incidence of GDM was 9.4%.

Despite the differences in settings, selective screening using a risk scoring algorithm nonetheless seemed to be effective.

6.3 STUDY STRENGTHS AND LIMITATIONS

6.3.1 Strengths of study

Firstly, all three studies of this PhD research addressed research questions that have not been completely answered before. The previous systematic reviews on universal versus selective screening were limited in terms of the databases searched (Hieronimus & Le Meaux, 2010), the types of studies included (Tieu *et al.,* 2010), or the scope of the effectiveness of the studies (Scott *et al*., 2002; Waugh *et al*., 2010). The systematic review in this research provided a comprehensive synthesis of current evidence on the effectiveness and cost-effectiveness of universal versus selective screening for GDM. Griffiths *et al.* (2013) investigated the attitudes of pregnant women towards two-step GDM screening in Australia, while the present Q methodology study is a unique study that explores user perspectives towards the IADPSG one-step screening for GDM. Two studies carried out previously (Naylor *et al.,* 1997; Van Leeuwen *et al.,* 2010) assessed a risk score-based selective screening approach for two-step GDM screening, while this risk score study is distinct because it is the first study exploring the selective approach under the new IADPSG one-step GDM screening. Hence, the findings of this PhD research will make new contributions to the current knowledge.

Secondly, this PhD research used a mixed-method study design, involving a systematic review, a qualitative as well as a quantitative study. The Q methodology provided user perspectives that are considered to be essential to patient-centred

healthcare. The risk score quantitative study used robust research design and investigated a full list of potential risk factors for developing a risk scoring algorithm.

6.3.2 Limitations of study

The current review has followed the standard process of conducting a systematic review. Four main databases of Medline, EMBASE, Web of Science and Cochrane Database were used, and research literature in the English language was searched. One of the limitations was that some potentially relevant studies, which were not reported in English may have been left out.

The findings of the Q methodology study and the risk score study are limited in terms of generalisability. The two studies were conducted in a Chinese context, wherein the IADPSG one-step universal screening for GDM was adopted. Due to the physical and cultural differences in population characteristics and differences in methods of screening tests, the conclusions of the two studies can be claimed to be applicable only to the Chinese population. However, the methods are generalisable, and they provide an important reference for other countries, especially those countries which are also implementing the IADPSG one-step screening approach.

The risk scoring algorithm was established and assessed using a retrospective cohort of 2897 pregnant women who gave birth to their babies at the hospital in 2013 (derivation cohort). A further validation cohort to validate the effectiveness of the risk scoring algorithm is needed, preferably a cohort from other areas of China. A user-friendly computer or web-based tool, similar to the QRISK2 tool for cardiovascular risk calculation (https://www.qrisk.org/2016/) might be needed in the future to facilitate calculation and enhance ease of application.

6.4 STUDY IMPLICATIONS

6.4.1 Implications for practice

The systematic review has analysed and drawn conclusions based on existing evidence in the field. Universal screening is recommended for countries or areas wherein GDM prevalence is relatively high and economic constraints are not experienced. For areas where GDM prevalence is low, it is recommended that they retain their current practice (whether it is universal or selective screening) before more robust evidence emerges. For countries which are implementing selective screening, a risk score-based selective screening approach could be explored. For the risk score calculation in real practice during future implementation, it is easy for practitioners to make calculations when they are facilitated by a simple web- or computer-based calculation tool. A similar tool is the QRISK2 tool for cardiovascular risk calculation (https://www.qrisk.org/2016/).

Pregnant women generally believe that the IADPSG one-step screening approach is important and necessary to be carried out on all expectant mothers. However, the non-GDM women feel strongly that the OGTT test is inconvenient and a burden. Also both the GDM and non-GDM participants felt that they would like more information both before and after the OGTT. In GDM screening services, it is recommended that a detailed GDM and OGTT information leaflet should be offered to pregnant women both before and after the OGTT to meet their health needs.

Translating the research evidence to real-world guidelines, policymaking and implementation can be challenging and time-consuming. While a great deal of research evidence exists, relatively little has been disseminated, taken up or applied in practice. Currently, as described in the Chapter 1, substantial inconsistencies and

controversies exist between research evidence, guidelines and different field practices for gestational diabetes screening. The research evidence is of little relevance if it remains unused by health clinicians and policymakers.

However, policymaking or guideline setting is a process of deliberation. Apart from scientific evidence, other factors such as values and the culture of each country can be influential. Countries have different values in relation to effectiveness and cost-effectiveness, which are sometimes associated with, but not always relevant to, the economic level of the country. For example, cost-effectiveness of heath care is emphasised in the UK (Raspe, 2016), which might be one of the reasons for implementation of selective screening in the UK as suggested by NICE (2015). Secondly, different countries have different normative cultures in healthcare. For example, Sweden and Norway have person-centred solidarity, whereas the UK has community-centred solidarity (Raspe, 2016). The UK tends to care more about health maximisation as a whole, compared to the person-centred solidarity, which tends to prioritise the worst off or sickest. Under such cultural imperatives, the UK presents a large chance to balance the benefit of avoiding unnecessary OGTT/burden for non-GDM women (over 90% of all pregnant women) and the cost of missing a small proportion of GDM women. Thirdly, understanding patient needs and promoting patient involvement in healthcare decision making is being increasingly emphasised and implemented. Therefore, it is considered of value to provide space for the projection of patients' voices when making any healthcare decision including GDM screening.

6.4.2 Recommendations for future research

The systematic review identified only one cost-effectiveness study on universal versus selective screening for GDM (Poncet *et al*., 2002). This showed a slight

difference between, indicating that the cost to obtain one unit of additional effectiveness under universal screening was 1.1 times more expensive than that for selective screening. More cost-effectiveness studies will be needed to advance more robust conclusions.

There are only three existing studies on a risk score-based selective screening approach for GDM, each demonstrating that developing a risk-scoring algorithm based on the local population profile made selective screening an effective screening approach. However, more risk score studies are recommended for each country. The effectiveness of the risk score-based selective screening approach is high dependent on the local setting of each country. When conducting future risk-score studies, it is essential that the researchers should investigate a full range of potential GDM risk factors to maximise the accuracy and effectiveness of the prediction model, since it is uncertain whether any other simple biomarkers can improve upon this prediction. A recent study (Meek *et al*., 2016) showed how even random glucose is a better predictor of GDM than BMI or maternal age during the first trimester. Some recent studies showed that the homeostatic model assessment for insulin resistance index (HOMA IR) level was associated with GDM and could be used as a predictor (Mohamed *et al*., 2013; Alptekin *et al*., 2016). Whether these hold true in Chinese population is a matter of research in the future. Additionally, it also remains to be established whether the random glucose indicator could be improved by using HbA1c as a predictor, either on its own or as a composite risk marker. Whether any urinary metabolomics during 1st trimester could be used as predictor for GDM worth further investigation. In the study of Sachse *et al.* (2012), an increase of excreted urinary citrate correlated with the severity of GDM was observed. Meta-analysis of candidate gene studies and genome-wide association analysis (GWAS) have identified a number of genes which were reproducibly associated with GDM, including *TCF7L2, GCK, KCNJ11, KCNQ1, CDKAL1, IGF2BP2, MTNR1B,* and

IRS1 (Lowe *et al*., 2016). These genes are also associated with T2DM. Genetics of T2DM and GDM as predictive markers of GDM are also worth exploring in the future.

Since only two studies on pregnant women's perspectives on GDM screening have been conducted in Australia (Griffiths *et al.*, 1993) and China, with the Australian study being conducted early in 1993, future qualitative perspective studies are needed under other local screening settings other than China. This is to ensure that patient needs and viewpoints are understood fully so as to facilitate decisionmaking for each country.

There is trend to transfer the outcome measure of test accuracy to clinical outcomes for screening studies. The systematic review synthesised the test accuracy of selective screening in comparison with universal screening. As explained in the Discussion section in Chapter 3, the present systematic review did not synthesise clinical outcomes is because there were only three studies (Griffin *et al.*, 2000; Cosson *et al.*, 2006; Ezimokhai *et al*., 2006) using the clinical outcome measures but each had the same bias in study design. The most appropriate study design compares the clinical outcomes of all pregnant women not just GDM women under the two screening approaches. Therefore, it is recommended that future studies are designed in line with this consideration. Moreover, future studies can also examine the clinical outcomes of the GDM women with lowest risk. These are the GDM women who would be missed during the application of selective screening as having none or low risks; whether or not these lowest risk GDM women develop adverse outcomes has important implications for assessing the selective screening approach. As a further step beyond using clinical outcome measures, future studies could consider using a decision tree to illustrate the two choices of universal screening and selective screening approaches. A decision tree is a decision support tool that uses a tree-like

graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Use of this technique can further facilitate the judgment and decision making in clinical medicine.

6.5 CONCLUSIONS

The optimal GDM screening approach is ‘setting dependent’. It is influenced by the population characteristics and risks, the different screening tests and criteria used, and the patients’ perspectives, as well as wider context of healthcare value and culture. This PhD research has provided scientific evidence on the effectiveness and cost-effectiveness of universal versus selective screening and patient perspectives, and it has also explored a risk score-based selective screening approach under the new IADPSG criteria.

Although it appears that universal screening is the best approach, there are a number of associated issues pertaining to the cost of resources, patient burden and lack of cost-effectiveness. This means that using the universal approach to identify and manage all pregnant women may not really be worth it, especially for those at the lower end of the risk. Selective screening is therefore believed to be the better approach, but how this can be conducted is the most pertinent question. Ideally each population should have its own high risk screening criterion. Potentially, selective screening with criteria that is appropriate for the population can be efficient; by not employing universal screening can save numerous costs and the anxiety of undergoing OGTT. Nevertheless, the cost of the missing GDM and their long-term consequences after delivery through selective screening needs to be further explored by future studies.

The Q methodology study highlights the burden and inconvenience around OGTT and the diagnosis of GDM. There needs to be further work to minimise this by better education as well as selective screening of individuals with varying risks. The case-control study has shown weighting of simple risk factors can be a better predictor of risk. Using the risk score algorithm could spare 41% of pregnant women from screening, while accurately identifying 80% GDM cases. Whether one hundred percent of the cases can be picked up is the debate. However, even if 20% GDM cases are missed (as in this study), these are the GDM women with the lowest preexisting risk factor profile. Further prospective study is ideally needed to examine the clinical outcomes of these lowest risk GDM women, so as to justify the adoption of the selective screening approach.

In terms of real-world implementation, aside from the scientific evidence on effectiveness, cost-effectiveness and patient perspectives investigated by this research, it needs to be borne in mind that the country's value and culture in healthcare policy can also contribute to final healthcare decision making. It needs to be noted that such decision making is a deliberate and time-consuming process.

This thesis has successfully addressed the research aims and the stated research objectives. With regard to further investigation into the optimal screening approach for GDM, several other important areas for future research have been identified. More cost-effectiveness studies on selective versus universal screening are needed. Testing the effectiveness of risk score-based selective screening is highly recommended for other countries. This is especially true for countries that have not assessed the effectiveness of a selective screening approach and for countries wishing to explore the possibility of improving their selective screening approach. A full range of traditional and novel biomarkers should be investigated when developing the selection criteria for identifying high risk women. Within literature,

there is a clear gap for more country-specific research on understanding pregnant women's perspectives on GDM screening in their own setting. Furthermore, more screening studies using clinical outcome measures with the correct study design and more prognostic studies are needed. It is believed that the findings of future research in light of the recommendations in this study will offer stronger and more specific evidence for making a robust evaluation of the best screening approach for each country.

APPENDICES

Appendix 1. Search strategy in Medline, EMBASE, Cochrane Database, and Web of Science.

Appendix 1.1 Search strategy in Medline (searched on 1April 2013 and updated on 14 November 2014)

	Searches	Results	Search Type
1	exp Mass Screening/	105316	Advanced
2	"screen*".m_titl.	115146	Advanced
3	exp Diabetes, Gestational/	7494	Advanced
4	(diabet* adj3 gestation*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept, rare disease supplementary concept, unique identifier]	8855	Advanced
5	gdm.mp.	2876	Advanced
6	1 or 2	170957	Advanced
7	3 or 4 or 5	10221	Advanced
8	6 and 7	894	Advanced
9	**limit 8 to (English language and yr="1980 -Current")**	**767**	**Advanced**

Appendix 1.2 Search strategy in Embase (searched on 1 April 2013 and updated on 14 November 2014)

	Searches	Results	Search Type
1	exp mass screening/ or exp screening/	485964	Advanced
2	"screen*".m_titl.	152843	Advanced
3	exp pregnancy diabetes mellitus/	21701	Advanced

4	(diabet* adj3 gestation*).mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]	13846	Advanced
5	gdm.mp.	5168	Advanced
6	1 or 2	529667	Advanced
7	3 or 4 or 5	24866	Advanced
8	6 and 7	2444	Advanced
9	**limit 8 to (English language and yr="1980 -Current")**	**2154**	**Advanced**

Appendix 1.3 Search strategy in SCI and SSCI (Web of Science) (searched on 1 April 2013 and updated on 14 November 2014)

Searched on Web of Science on (gestation* same diabet*) and screen* in 'TITLE' from 1980 onwards.

Results
568 articles

Appendix 1.4 Search strategy in Cochrane Database (searched on 1 April 2013 and updated on 14 November 2014)

Searched on (gestation* near diabet*) and screen*.

All Results (81)
Cochrane Reviews (3)
Other Reviews (8)
Trials (55)
Methods Studies (0)
Technology Assessments (9)
Economic Evaluations (13
Cochrane Groups (0)

Appendix 2. Table of included studies for systematic review

Appendix 2.1 Effectiveness studies

	Reference
	Effectiveness study with two-step GDM tests
1	Arora, D., Arora, R., Sangthong, S., Leelaporn, W. & Sangratanathongchai, J. (2013) Universal screening of gestational diabetes mellitus: prevalence and diagnostic value of clinical risk factors. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet*, 96 (3): 266-271.
2	Caliskan, E., Kayikcioglu, F., Ozturk, N., Koc, S. & Haberal, A. (2004) A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. *Acta Obstetricia et Gynecologica Scandinavica*, 83 (6): 524-530.
3	Cosson, E., Benbara, A., Pharisien, I., Nguyen, M. T., Revaux, A., Lormeau, B., Sandre-Banon, D., Assad, N., Pillegand, C., Valensi, P. & Carbillon, L. (2013) Diagnostic and prognostic performances over 9 years of a selective screening strategy for gestational diabetes mellitus in a cohort of 18,775 subjects. *Diabetes Care*, 36 (3): 598-603.
4	Coustan, D. R., Nelson, C., Carpenter, M. W., Carr, S. R., Rotondo, L. & Widness, J. A. (1989) Maternal Age And Screening For Gestational Diabetes - A Population-Based Study. *Obstetrics and Gynecology*, 73 (4): 557-561.
5	Danilenko-Dixon, D. R., Van Winter, J. T., Nelson, R. L. & Ogburn Jr, P. L. (1999) Universal versus selective gestational diabetes screening: Application of 1997 American Diabetes Association recommendations. *American Journal of Obstetrics and Gynecology*, 181 (4): 798-802.
6	Di Cianni, G., Volpe, L., Lencioni, C., Miccoli, R., Cuccuru, I., Ghio, A., Chatzianagnostou, K., Bottone, P., Teti, G., Del Prato, S. & Benzi, L. (2003) Prevalence and risk factors for gestational diabetes assessed by universal screening. *Diabetes Research and Clinical Practice*, 62 (2): 131-137.
7	Hadaegh, F., Tohidi, M., Harati, H., Kheirandish, M. & Rahimi, S. (2005) Prevalence of gestational diabetes mellitus in southern Iran (Bandar Abbas City). *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*, 11 (5): 313-318.
8	Helton, M. R., Arndt, J., Kebede, M. & King, M. (1997) Do low-risk prenatal patients really need a screening glucose challenge test? *Journal of Family Practice*, 44 (6): 556-561.
9	Jensen, D. M., Molsted-Pedersen, L., Beck-Nielsen, H., Westergaard, J. G., Ovesen, P. & Damm, P. (2003) Screening for gestational diabetes mellitus by a model based on risk indicators: A prospective study. *American Journal of Obstetrics and Gynecology*, 189 (5): 1383-1388.
10	Jimenez-Moleon, J. J., Bueno-Cavanillas, A., Luna-del-Castillo, J. D., Garcia-Martin, M., Lardelli-Claret, P. & Galvez-Vargas, R. (2002) Prevalence of gestational diabetes mellitus: variations related to screening strategy used. *European Journal of Endocrinology*, 146 (6): 831-837.

11	Lavin Jr, J. P. (1985) Screening of high-risk and general populations for gestational diabetes. Clinical application and cost analysis. *Diabetes*, 34 (SUPPL. 2): 24-27.
12	Moses, R., Griffiths, R. & Davis, W. (1995) Gestational diabetes: do all women need to be tested? *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 35 (4): 387-389.
13	Moses, R. G., Moses, J. & Davis, W. S. (1998) Gestational diabetes: Do lean young caucasian women need to be tested? *Diabetes Care*, 21 (11): 1803-1806.
14	Naylor, C. D., Sermer, M., Chen, E. & Farine, D. (1997) Selective screening for gestational diabetes mellitus. *New England Journal of Medicine*, 337 (22): 1591-1596.
15	Sacks, D. A., Abu-Fadil, S., Karten, G. J., Forsythe, A. B. & Hackett, J. R. (1987) Screening for gestational diabetes with the one-hour 50-g glucose test. *Obstetrics & Gynecology*, 70 (1): 89-93.
16	Teh, W. T., Teede, H. J., Paul, E., Harrison, C. L., Wallace, E. M. & Allan, C. (2011) Risk factors for gestational diabetes mellitus: Implications for the application of screening guidelines. *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 51 (1): 26-30.
17	Van Leeuwen, M., Opmeer, B. C., Zweers, E. J. K., Van Ballegooie, E., Ter Brugge, H. G., De Valk, H. W., Visser, G. H. A. & Mol, B. W. J. (2010) Estimating the risk of gestational diabetes mellitus: A clinical prediction model based on patient characteristics and medical history. *BJOG: An International Journal of Obstetrics and Gynaecology*, 117 (1): 69-75.
18	Williams, C. B., Iqbal, S., Zawacki, C. M., Yu, D., Brown, M. B. & Herman, W. H. (1999) Effect of selective screening for gestational diabetes. *Diabetes Care*, 22 (3): 418-421.
19	Zoller, D. P., Jurica, J. V., Gould, S. H. & Weinstein-Mayer, S. (1988) Screening for gestational diabetes. *Journal of the American Board of Family Practice*, 1 (2): 98-100.
	Effectiveness study with a one-step GDM test
20	Capula, C., E. Chiefari, *et al.* (2013). "Gestational diabetes mellitus: Screening and outcomes in southern Italian pregnant women." *ISRN Endocrinology* 1(1).
21	Chong, Y. S., S. Cai, *et al.* (2014). "Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: A cohort study." BMC pregnancy and childbirth 14(1).
22	Corrado, F., B. Pintaudi, *et al.* (2014). "Italian risk factor-based screening for gestational diabetes." Journal of Maternal-Fetal and Neonatal Medicine 27(14): 1445-1448
23	Ostlund, I. & Hanson, U. (2003) Occurrence of gestational diabetes mellitus and the value of different screening indicators for the oral glucose tolerance test. *Acta Obstetricia Et Gynecologica Scandinavica*, 82 (2): 103-108.
24	Pintaudi, B., G. Di Vieste, *et al.* (2014). "Improvement of selective screening strategy for gestational diabetes through a more accurate definition of high-risk groups." European Journal of Endocrinology 170(1): 87-93.

25	Savona-Ventura, C., Vassallo, J., Marre, M. & Karamanos, B. G. (2013) A composite risk assessment model to screen for gestational diabetes mellitus among Mediterranean women. *International Journal of Gynecology and Obstetrics*, 120 (3): 240-244.
26	Shamsuddin, K., Mahdy, Z. A., Rafiaah, I. S., Jamil, M. A. & Rahimah, M. D. (2001) Risk factor screening for abnormal glucose tolerance in pregnancy. *International Journal of Gynecology and Obstetrics*, 75 (1): 27-32.
27	Shirazian, N., Emdadi, R., Mahboubi, M., Motevallian, A., Fazel-Sarjuei, Z., Sedighpour, N., Fadaki, S. F. & Shahmoradi, N. (2009) Screening for gestational diabetes: usefulness of clinical risk factors. *Archives of Gynecology and Obstetrics*, 280 (6): 933-937.
28	Wagaarachchi, P. T., Fernando, L., Premachadra, P. & Fernando, D. J. S. (2001) Screening based on risk factors for gestational diabetes in an Asian population. *Journal of Obstetrics and Gynaecology*, 21 (1): 32-34.

Appendix 2.2 Cost-effectiveness and cost studies

Study	Reference
	Cost-effectiveness study
29	Poncet, B., Touzet, S., Rocher, L., Berland, M., Orgiazzi, J. & Colin, C. (2002) Cost-effectiveness analysis of gestational diabetes mellitus screening in France. *European Journal of Obstetrics, Gynecology, & Reproductive Biology*, 103 (2): 122-129.
	Cost study
30	Coustan, D. R., Nelson, C., Carpenter, M. W., Carr, S. R., Rotondo, L. & Widness, J. A. (1989) Maternal Age And Screening For Gestational Diabetes - A Population-Based Study. Obstetrics and Gynecology, 73 (4): 557-561.
31	Larijani, B., Hossein-Nezhad, A. & Vassigh, A. R. (2004) Effect of varying threshold and selective versus universal strategies on the cost in gestational diabetes mellitus. *Archives of Iranian Medicine*, 7 (4): 267-271.
32	Reed, B. D. (1984) Screening for gestational diabetes--analysis by screening criteria. *Journal of Family Practice*, 19 (6): 751-755.
33	Shamsuddin, K., Mahdy, Z. A., Rafiaah, I. S., Jamil, M. A. & Rahimah, M. D. (2001) Risk factor screening for abnormal glucose tolerance in pregnancy. *International Journal of Gynecology and Obstetrics*, 75 (1): 27-32.

Appendix 3. Table of excluded studies and reasons for exclusion

	Excluded studies	Reason for exclusion
1	Agarwal, M. M. & Punnose, J. (2001) Screening for gestational diabetes in high-risk populations: The United Arab Emirates experience. *Annals of Saudi Medicine*, 21 (1-2): 117-119.	Comparing different thresholds of GDM screening test
2	Al-Bassam, M. K. S., Al-Awar, S., Khan, F., Karim, Q. A., Al-Shibli, A. I. & Chedid, F. (2007) Universal screening strategy for gestational diabetes mellitus: The experience of Tawam Hospital. *Emirates Medical Journal*, 25 (3): 307-310.	Outcome measures only included sensitivity, but no specificity
3	Alberico, S., Strazzanti, C., De Santo, D., De Seta, F., Lenardon, P., Bernardon, M., Zicari, S. & Guaschino, S. (2004) Gestational diabetes: Universal or selective screening? *Journal of Maternal-Fetal and Neonatal Medicine*, 16 (6): 331-337.	Outcome measures only included sensitivity, but no specificity
4	Arcidiacono, B, Capula, C, Chiefari, E, Ventura, V, Iiritano, S, Vero, A, Puccio, L, Pullano, V, Foti, D, Brunetti, A, & Vero, R n.d., (2013). "Gestational diabetes screening: New considerations on the recent Italian recommendations." *Diabetologia* 56: S512.	Abstract only
5	Avalos, G. E., Owens, L. & Dunne, F. (2012) How many women with gestational diabetes mellitus are missed if selective screening strategies are used? *Diabetologia*, 55 S446.	Conference publication with abstract only
6	Bachaoui, M., K. Benharrat, *et al.* (2014). "Gestational diabetes: What is the impact of the adoption of the criteria of IADPSG?" *Diabetologia* 1): S445-S446.	Abstract only
7	Baliutaviciene, D., Petrenko, V. & Zalinkevicius, R. (2002) Selective or universal diagnostic testing for gestational diabetes mellitus. *International Journal of Gynaecology & Obstetrics*, 78 (3): 207-211.	Outcome measures only included sensitivity, but no specificity
8	Bassaw, B., Mohammed, N., Ramsewak, S., Bassawh, L., Khan, A., Bhola, M. & Chekuri, A. (2012) Pregnancy outcome among women universally screened for gestational diabetes mellitus with a lime-flavoured drink. *Journal of Obstetrics and Gynaecology*, 32 (5): 422-425.	Outcome measures only included sensitivity, but no specificity
9	Beard, R. W., Gillmer, M. D. G., Oakley, N. W. & Gunn, P. J. (1980) SCREENING FOR GESTATIONAL DIABETES. *Diabetes Care*, 3 (3): 468-471.	Literature review
10	Bell, R., Hayes, L., Lewis-Barned, N., Bilous, M., Brandon, H., Pearson, S., Emmerson, C., Adair, S., Crowder, D. & Bilous, R. (2010) Diagnosis, treatment and outcome of gestational diabetes: A multi-centre study in north-east England (NorGES). *Diabetic Medicine*, 1) 15.	Conference publication with abstract only

11	Berger, H., Crane, J., Farine, D., Armson, A., De La Ronde, S., Keenan-Lindsay, L., Leduc, L., Reid, G. & Van Aerde, J. (2002) Screening for gestational diabetes mellitus. *Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada* : *JOGC*, 24 (11): 894-912.	Literature review
12	Brody, S. C., Harris, R. & Lohr, K. (2003) Screening for gestational diabetes: A summary of the evidence for the U.S. Preventive Services Task Force. *Obstetrics and Gynecology*, 101 (2): 380-392.	Systematic review (and didn't include universal versus selective screening)
13	Calonge, N., Petitti, D. B., DeWitt, T. G., Gordis, L., Gregory, K. D., Harris, R., Isham, G., LeFevre, M. L., Loveland-Cherry, C., Marion, L. N., Moyer, V. A., Ockene, J. K., Sawaya, G. F., Siu, A. L., Teutsch, S. M. & Yawn, B. P. (2008) Screening for gestational diabetes mellitus: U.S. preventive services task force recommendation statement. *Annals of Internal Medicine*, 148 (10): 759-765.	Systematic review (and didn't include universal versus selective screening)
14	Cao, X. P., Chen, S. J., Wang, Z. L., Chen, S., Li, Y. B. & Xiao, H. P. (2009) The Value of Screening Based on Traditional Risk Factors Is Low in Identifying Gestational Diabetes and High in Predicting Postpartum Diabetes. *Diabetes*, 58 A470-A470.	Outcome measures didn't include sensitivity and specificity of screening; conference publication with abstract only
15	Chen, P. Y., E. A. Finkelstein, *et al.* (2014). "Cost-effectiveness analysis on gestational diabetes mellitus screening strategies." *Journal of Maternal-Fetal and Neonatal Medicine* 27: 69-70.	Abstract only. Full text not available.
16	Chevalier, N., Hieronimus, S., Giaume, V., Brucker-Davis, F., Bongain, A. & Fenichel, P. (2009) Obstetrical outcomes in pregnancies with gestational diabetes: what benefits? Which patients? *Diabetologia*, 52 (S1) S458.	Abstract only
17	Chevalier, N., Fenichel, P., Giaume, V., Loizeau, S., Bongain, A., Daideri, G., Brucker-Davis, F. & Hieronimus, S. (2011) Universal two-step screening strategy for gestational diabetes has weak relevance in French Mediterranean women: Should we simplify the screening strategy for gestational diabetes in France? *Diabetes & Metabolism*, 37 (5): 419-425.	Outcome measures only included sensitivity, but no specificity
18	Corcoy, R., Garcia-Patterson, A., Pau, E., Pascual, E., Altirriba, O., Adelantado, J. M. & de Leiva, A. (2004) Is selective screening for gestational diabetes mellitus worthwhile everywhere? *Acta Diabetologica*, 41 (4): 154-157.	The specificity couldn't be calculated using the definition in the review
19	Cosson, E., Benchimol, M., Carbillon, L., Pharisien, I., Paries, J., Valensi, P., Lormeau, B., Bolie, S., Uzan, M. & Attali, J. R. (2006) Universal rather than selective screening for gestational diabetes mellitus may improve fetal outcomes. *Diabetes & Metabolism*, 32 (2): 140-146.	Outcome measures only included sensitivity, but no specificity

20	Cosson, E., C. Cussac-Pillegand, *et al.* (2014). "The diagnostic and prognostic performance of a selective screening strategy for gestational diabetes mellitus according to ethnicity in Europe." *Journal of Clinical Endocrinology & Metabolism* 99(3): 996-1005.	Duplicate with an included study (Cosson *et al*., 2013)
21	Davey, R. X. & Hamblin, P. S. (2001) Selective versus universal screening for gestational diabetes mellitus: an evaluation of predictive risk factors. *Medical Journal of Australia*, 174 (3): 118-121.	Outcome measures included sensitivity and specificity. However, sensivitity was calulated based on all risk factors, while specificity was calculated only based on age.
22	Di Cianni, G., Volpe, L., Casadidio, I., Bottone, P., Marselli, L., Lencioni, C., Boldrini, A., Teti, G., Del Prato, S. & Benzi, L. (2002) Universal screening and intensive metabolic management of gestational diabetes: cost-effectiveness in Italy. *Acta Diabetologica*, 39 (2): 69-73.	Universal screening approah included screening and diagnosis test, however, selective screening approach only included diagnosis test. Not comparatable.
23	Dietrich, M. L., Dolnicek, T. F. & Rayburn, W. F. (1987) GESTATIONAL DIABETES SCREENING IN A PRIVATE, MIDWESTERN AMERICAN POPULATION. *American Journal of Obstetrics and Gynecology*, 156 (6): 1403-1408.	The specificity couldn't be calculated using the definition in the review
24	Donovan, L., Hartling, L., Muise, M., Guthrie, A., Vandermeer, B. & Dryden, D. M. (2013) Screening Tests for Gestational Diabetes: A Systematic Review for the US Preventive Services Task Force. *Annals of internal medicine*,	Systematic review (and didn't include universal versus selective screening)
25	Ezimokhai, M., Joseph, A. & Bradley-Watson, P. (2006) Audit of pregnancies complicated by diabetes from one center five years apart with selective versus universal screening. *Annals of the New York Academy of Sciences*, 1084 132-140.	Outcome measures didn't include sensitivity and specificity of screening
26	Farrar, D., Wright, J., Whitelaw, D. & Tuffnell, D. (2011) Evaluation of the impact of universal screening for gestational diabetes mellitus on maternal and neonatal health outcomes. *Archives of Disease in Childhood: Fetal and Neonatal Edition*, 96 Fa126.	Outcome measures didn't include sensitivity and specificity of screening; conference publication with abstract only

27	Griffin, M. E., Coffey, M., Johnson, H., Scanlon, P., Foley, M., Stronge, J., O'Meara, N. M. & Firth, R. G. (2000) Universal vs. risk factor-based screening for gestational diabetes mellitus: detection rates, gestation at diagnosis and outcome. *Diabetic Medicine*, 17 (1): 26-32.	Outcome measures didn't include sensitivity and specificity of screening
28	Hall, C., Going, A., Moutter, S., Thynne, A. D., Salloum, M., Sengupta, S. & Cummings, M. H. (2009) Highlighting the dilemmas of GTTs performed during pregnancy: Are we screening appropriately? *Diabetic Medicine*, 26 178.	Conference publication with abstract only
29	Hapuarachi, S., N. Rahman, *et al.* (2014). "Is national targeted screening for gestational diabetes fit for purpose? Experience from East of England hospital." BJOG: An International Journal of Obstetrics and Gynaecology 121: 161.	Abstract only
30	Healy, G. M., Vellinga, A., Carmody, L., Avalos, G., Mustafa, E., Khalil, S., Noctor, E., Kirwan, B. & Dunne, F. P. (2012a) Atlantic DIP: Universal vs. Selective Screening for Gestational Diabetes (GDM). *Diabetes*, 61 A641.	Outcome measures didn't include sensitivity and specificity of screening
31	Healy, G. M., Vellinga, A., Carmody, L., Avalos, G., Mustafa, E., Khalil, S., Noctor, E., Kirwan, B. & Dunne, F. P. (2012b) Atlantic DIP: What does selective screening for gestational diabetes miss? *Diabetologia*, 55 S38.	Conference publication with abstract only
32	Hieronimus, S. & Le Meaux, J. P. (2010) Relevance of gestational diabetes mellitus screening and comparison of selective with universal strategies. *Diabetes & Metabolism*, 36 (6): 575-586.	Systematic review (included 14 studies on univesal versus selective screening)
33	Hillier, T. A., Vesco, K. K., Pedula, K. L., Beil, T. L., Whitlock, E. P. & Pettitt, D. J. (2008) Screening for gestational diabetes mellitus: a systematic review for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, 148 (10): 766-775.	Systematic review (and didn't include universal versus selective screening)
34	Jang, H. C. (1995) Screening for gestational diabetes mellitus in Korea. *International Journal of Gynecology and Obstetrics*, 51 (2): 115-122.	Outcome measures didn't include sensitivity and specificity of screening
35	Jimenez-Moleon, J. J., Bueno-Cavanillas, A., Luna-del-Castillo, J. D., Lardelli-Claret, P., Garcia-Martin, M. & Galvez-Vargas, R. (2000) Predictive value of a screen for gestational diabetes mellitus: influence of associated risk factors. *Acta Obstetricia Et Gynecologica Scandinavica*, 79 (11): 991-998.	Same with the included study of Jimenez-Moleon*et al.* (2002)

36	Lacaria, E., C. Lencioni, *et al.* (2014). "National guidelines for gestational diabetes mellitus (GDM) screening in Italy: Application and effectiveness." *Diabetologia* 1): S446.	Abstract only
37	Lacey, A., Roche, J. & Wheatley, T. (2011) Screening for gestational diabetes: Are NICE risk factors adequate? *Diabetic Medicine*, 28 182-183.	Conference publication with abstract only
38	Lappin, S. M., Watt, P., Traub, A. I., Tharma, S., Courtney, H. & McCance, D. R. (2010) Audit of risk factors for Gestational Diabetes (GDM) in women diagnosed by a universal screening programme. *Diabetic Medicine*, 1) 173-174.	Conference publication with abstract only
39	Lindqvist, M., M. Persson, *et al.* (2014). "No consensus on gestational diabetes mellitus screening regimes in Sweden: Pregnancy outcomes in relation to different screening regimes 2011 to 2012, a cross-sectional study." BMC pregnancy and childbirth 14(1).	The specificity couldn't be calculated using the definition in the review
40	Lohse, N., Marseille, E. & Kahn, J. G. (2011) Development of a model to assess the cost-effectiveness of gestational diabetes mellitus screening and lifestyle change for the prevention of type 2 diabetes mellitus. *International Journal of Gynecology & Obstetrics*, 115 S20-S25.	Cost-effectiveness study, but didn't include universal versus selective screening
41	Magenheim, R., Schafer-Graf, U., Supak, D., Feher, Z., Abou-Dakn, M. & Tamas, G. (2010) Early gestational diabetes mellitus (GDM) screening based on risk factors. *Archives of Gynecology and Obstetrics*, 282 S154-S155.	Conference publication with abstract only
42	Marquette, G. P., Klein, V. R. & Niebyl, J. R. (1985) Efficacy of screening for gestational diabetes. *American Journal of Perinatology*, 2 (1): 7-9.	Full text is not available
43	Mazze, R. S. & Krogh, C. L. (1992) Gestational diabetes mellitus: now is the time for detection and treatment. *Mayo Clinic Proceedings*, 67 (10): 995-1002.	Did not compare universal screening with selective screening
44	Mettayil, J. & Marshall, S. (2009) Gestational diabetes: Screening criteria. *Diabetic Medicine*, 26 83-84.	Conference publication with abstract only
45	McCarthy, A. D., Curciarello, R., Castiglione, N., Tayeldin, M. F., Costa, D., Arnol, V., Prospitti, A., Aliano, A., Archuby, D., Graieb, A., Torres, M. J., Etcheverry, S. B. & Apezteguia, M. C. (2010) Universal versus selective screening for the detection, control and prognosis of gestational diabetes mellitus in Argentina. *Acta Diabetologica*, 47 (2): 97-103.	Outcome measures didn't include sensitivity and specificity of screening
46	Minsart, A. F., Lescrainier, J. P. & Vokaer, A. (2009) Selective versus Universal Screening for Gestational Diabetes Mellitus: An Evaluation of Naylor's Model. *Gynecologic and Obstetric Investigation*, 68 (3): 154-	The specificity couldn't be calculated using the definition in

	159.	the review
47	Neelakandan, R. and P. Shankar Sethu (2014). "Early universal screening for gestational diabetes mellitus." Journal of Clinical and Diagnostic Research 8(4): OC12-OC14.	Neither specificity nor sensitivity was reported
48	Nicholson, W. K., Fleisher, L. A., Fox, H. E. & Powe, N. R. (2005) Screening for gestational diabetes mellitus: A decision and cost-effectiveness analysis of four screening strategies. *Diabetes Care*, 28 (6): 1482-1484.	Cost-effectiveness study, but didn't include universal versus selective screening
49	Nijjar, S. K., Hunt, K. F., Rogers, H., Smith, C., Gayle, C. M., Marsh, M. S., Amiel, S. A. & Choudhary, P. (2011) Clinical outcomes of patients with gestational diabetes mellitus who do not have typical risk factors. *Diabetologia*, 54 S479-S480.	Conference publication with abstract only
50	Phaloprakarn, C., Tangjitgamol, S. & Manusirivithaya, S. (2009) A risk score for selective screening for gestational diabetes mellitus. *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 49 (5): 572-572.	Intervention and comparator didn't include GDM diagnosis test
51	Poyhonen-Alho, M. K., Teramo, K. A., Kaaja, R. J. & Hiilesmaa, V. K. (2005) 50gram oral glucose challenge test combined with risk factor-based screening for gestational diabetes. *European journal of obstetrics, gynecology, and reproductive biology*, 121 (1): 34-37.	In the selective screening approach, risk factor selection was used to replace GCT screening test, but not used before GCT.
52	Ram, U., Gopal, J. & Mahadevan, S. (2009) An evaluation of the characteristics of gestational diabetes in an urban Indian setting. *International Journal of Gynecology and Obstetrics*, 107 S315.	Conference publication with abstract only
53	Russell, M. A., Carpenter, M. W. & Coustan, D. R. (2007) Screening and diagnosis of gestational diabetes mellitus. *Clinical Obstetrics and Gynecology*, 50 (4): 949-958.	Literature review
54	Scott, D. A., Loveman, E., McIntyre, L. & Waugh, N. (2002) Screening for gestational diabetes: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England)*, 6 (11): 1-161.	Literature review
55	Swinker, M. (1983) Routine screening for gestational diabetes mellitus in a family practice center. *The Journal of family practice*, 17 (4): 611.	Universal screening approah included screening and diagnosis test, however, selective screening approach only included

		diagnosis test. Not comparatable.
56	Tieu, J., Middleton, P., McPhee, A. J. & Crowther, C. A. (2010) Screening and subsequent management for gestational diabetes for improving maternal and infant health. *Cochrane Database of Systematic Reviews*, (7):	Systematic review (included one quasi-RCT on universal versus selective screening)
57	Van Leeuwen, M., Vijgen, S., Opmeer, B. C., Evers, I. & Mol, B. W. (2009) Cost-effectiveness analysis of screening for GDM. *American Journal of Obstetrics and Gynecology*, 1) S109.	Conference publication with abstract only
58	Waugh, N., Royle, P., Clar, C., Henderson, R., Cummins, E., Hadden, D., Lindsay, R. & Pearson, D. (2010) Screening for hyperglycaemia in pregnancy: a rapid update for the National Screening Committee. *Health Technology Assessment (Winchester, England)*, 14 (45): 1-183.	Systematic review (and didn't include universal versus selective screening)
59	Weeks, J. W., Major, C. A., de Veciana, M. & Morgan, M. A. (1994) Gestational diabetes: does the presence of risk factors influence perinatal outcome? *American Journal of Obstetrics & Gynecology*, 171 (4): 1003-1007.	Outcome measures only included sensitivity, but no specificity
60	Williams, C. B. & Herman, W. H. (1998) Universal vs. selective screening for gestational diabetes: A retrospective study. *Diabetes*, 47 A322-A322.	Abstract only (abstract not available)
61	Wilson, J. D. (2001) Gestational diabetes: universal or selective screening? Medical Journal of Australia, 174 (3): 113-114.	Abstract/full text not available
62	Wilson, N., Ashawesh, K., Smith, S. & Anwar, A. (2008) The cost of screening for gestational diabetes mellitus. *Journal of Medical Screening*, 15 (4): 213.	Only a summary of GDM screenig cost
63	Wylie, J., Walke, M., Shaw, N., Montague, I. & Millward, B. A. (2009) Screening for gestational diabetes: Audit of a 2007 cohort in light of NICE 63. *Diabetic Medicine*, 26 31.	Conference publication with abstract only
64	Yang, X. L., Hsu-Hage, B., Yu, L. C. & Simmons, D. (2002) Selective screening for gestational diabetes in Chinese women. *Diabetes Care*, 25 (4): 796-796.	Abstract only

Appendix 4 Quality assessment results

Appendix 4.1 Amended Downs and Black Checklist for assessing the effectiveness studies

The original Downs and Black checklist has 27 items with a maximum score of 32 points. The checklist was modified as not all items were relevant to the effectiveness studies in this review. All the 37 effectiveness studies were cohort studies/cross-sectional studies, no randomised controlled study was included. Item 23 and 24 about randomisation were removed. For all the cohort studies, selective screening and universal screening were from the same cohort, and the screening and diagnosis were conducted at the same time period. Item 21 and 22 about whether intervention and comparison group were from the same population and whether were recruited over the same period of time were removed. Since universal and selective screening group were form the same cohort, they were free from potential confounding factors. Item 5 and 25 about confounders were deleted. Sensitivity and specificity of selective screening compared to universal screening are main outcome measures. The two figures are fixed and invariable for each cohort. Item 7 about the estimate of variability for main outcomes and item 10 about the report of P-values were not applicable. The main adverse event of selective screening is missing GDM women, which was measured in terms of sensitivity. Item 8 about if all adverse events of intervention were reported will always be answered with Yes, thus was deleted. The screening and diagnosis of GDM are biomedical tests, which should not be affected by the blinding status of participants or the clinicians. Item 14 about blinding of study subjects and item 15 of blinding those measuring the outcomes was removed. Also, the biomedical tests results are automatically In addition, item 27 about power was revised from a scale of 0 to 5 to a scale of 0 to 1. It will be scored as 1 if there was power calculation in the study; and scored as 0 if there was no power calculation or explanation whether the number of participants was appropriate. The modified version of the checklist consists 16 items with a maximum score of 16 points. The checklist was employed for each of the effectiveness study of a cohort design. Higher scores reflected a superior quality of investigation.

Appendix 4.2 Quality assessment results of the effectiveness studies

Down & Black Checklist	**Arora *et al.* (2013)**	**Caliskan *et al.* (2004)**	**Capula *et al.* (2013)**	**Chong *et al.* (2014)**	**Corrado *et al.* (2014)**	**Cosson *et al.* (2013)**	**Coustan *et al.* (1989)**	**Danilenko-Dixon *et al.* (1999)**	**Di Cianni *et al.* (2003)**
1. Is the hypothesis/ aim/ objective of the study clearly described?	1	1	1	1	1	1	1	1	1
2.Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1	1	1	1	0	0	0	1	0
3.Are the characteristics of the patients included in the study clearly described ?	1	1	1	1	1	1	0	1	1
4. Are the interventions of interest clearly described?	1	1	1	1	1	1	1	1	1
6.Are the main findings of the study clearly described?	1	1	1	1	1	1	1	1	1
9.Have the characteristics of patients lost to follow-up been described?	1	1	1	1	1	0	1	1	1
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	1	1	1	1	1	1	1	1
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	1	1	1	1	0	1	1	1
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?	1	1	1	1	1	1	1	1	1
16.If any of the results of the study were based on"data dredging", was this made clear?	1	1	1	1	1	1	1	1	1
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	1	1	1	1	1	1	1	1	1
18. Were the statistical tests used to assess the	1	1	1	1	1	1	1	1	1

main outcomes appropriate?									
19.Was compliance with the intervention/s reliable?	1	1	1	1	1	1	1	1	1
20. Were the main outcome measures used accurate (valid and reliable)?	1	1	1	1	1	1	1	1	1
26.Were losses of patients to follow-up taken into account?	1	1	1	1	1	0	1	1	1
27.Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	0	0	0	0	0	0	0	0	0
Total score	15	16	15	15	14	11	13	15	14

NB: (1) 1-Yes; 0-No or unable to determine; (2) For the detailed description of how to judge each item, please see the original Downs and Black checklist (Downs & Black, 1998).

Down & Black Checklist	**Hadaegh *et al.* (2005)**	**Helton *et al.* (1997)**	**Jensen *et al.* (2003)**	**Jimenez-Moleon *et al.* (2002)**	**Lavin (1985)**	**Moses *et al.* (1995)**	**Moses *et al.* (1998)**	**Naylor *et al.* (1997)**	**Pintaudi *et al.* (2014)**
1. Is the hypothesis/ aim/ objective of the study clearly described?	1	1	1	1	1	1	1	1	1
2.Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1	1	1	1	0	1	1	1	1
3.Are the characteristics of the patients included in the study clearly described ?	1	0	1	1	0	1	1	1	1
4. Are the interventions of interest clearly described?	1	1	1	1	1	1	1	1	1
6.Are the main findings of the study clearly described?	1	1	1	1	1	1	1	1	1
9.Have the characteristics of patients lost to follow-up been described?	1	1	0	1	1	0	0	1	1
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	1	1	1	1	1	1	1	1
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	0	0	0	1	0	0	1	1
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?	1	1	1	1	1	1	1	1	1
16.If any of the results of the study were based on"data dredging", was this made clear?	1	1	1	1	1	1	1	1	1
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	1	1	1	1	1	1	1	1	1
18. Were the statistical tests used to assess the main outcomes appropriate?	1	1	1	1	1	1	1	1	1
19.Was compliance with the intervention/s reliable?	0	0	0	0	1	0	0	1	1

20. Were the main outcome measures used accurate (valid and reliable)?	1	1	1	1	1	1	1	1	1
26.Were losses of patients to follow-up taken into account?	1	1	1	1	1	0	0	1	1
27.Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	1	0	0	0	0	0	0	0	0
Total score	15	12	12	13	13	11	11	15	15

NB: (1) 1-Yes; 0-No or unable to determine; (2) For the detailed description of how to judge each item, please see the original Downs and Black checklist (Downs & Black, 1998).

Down & Black Checklist	**Sacks *et al.* (1987)**	**Teh *et al.* (2011)**	**Van Leeuwen *et al.* (2010)**	**Williams *et al.* (1999)**	**Zoller *et al.* (1988)**	**Ostlund & Hanson (2003)**	**Savona-Ventura *et al.* (2013)**	**Shamsuddin *et al.* (2001)**	**Shirazian *et al.* (2009)**	**Wagaarachchi *et al.* (2001)**
1. Is the hypothesis/ aim/ objective of the study clearly described?	1	1	1	1	1	1	1	1	1	1
2.Are the main outcomes to be measured clearly described in the Introduction or Methods section?	1	1	1	1	0	1	1	1	1	1
3.Are the characteristics of the patients included in the study clearly described ?	0	1	0	0	0	1	1	1	0	0
4. Are the interventions of interest clearly described?	1	1	1	1	1	1	1	1	1	1
6.Are the main findings of the study clearly described?	1	1	1	1	1	1	1	1	1	1
9.Have the characteristics of patients lost to follow-up been described?	1	1	0	0	1	1	1	0	1	1
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	1	1	1	1	1	1	0	1	1	1
12. Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	1	1	1	1	1	0	0	1	1	1
13. Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?	1	1	1	1	1	1	1	1	1	1
16.If any of the results of the study were based on"data dredging", was this made clear?	1	1	1	1	1	1	1	1	1	1
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	0	1	1	1	1	1	1	1	1	1
18. Were the statistical tests used to assess the main outcomes appropriate?	1	1	1	1	1	1	1	1	1	1
19.Was compliance with the intervention/s reliable?	1	1	0	1	1	1	1	1	1	1

20. Were the main outcome measures used accurate (valid and reliable)?	1	1	1	0	0	1	1	1	1	1
26.Were losses of patients to follow-up taken into account?	0	1	0	1	1	0	1	1	1	1
27.Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	0	0	0	0	0	0	0	0	0	0
Total score	12	15	11	12	12	13	13	14	14	14

NB: (1) 1-Yes; 0-No or unable to determine; (2) For the detailed description of how to judge each item, please see the original Downs and Black checklist (Downs & Black, 1998).

Appendix 4.3 Quality assessment results of the cost-effectiveness studies

Assessment result of Poncet *et al.*, (2002)

Item	Yes	No	Not clear	Not appropriate
Study design				
1. The research question is stated.	√	¨	¨	
2. The economic importance of the research question is stated.	√	¨	¨	
3. The viewpoint(s) of the analysis are clearly stated and justified.	√	¨	¨	
4. The rationale for choosing alternative programmes or interventions compared is stated.	¨	¨	√	
5. The alternatives being compared are clearly described.	√	¨	¨	
6. The form of economic evaluation used is stated.	√	¨	¨	
7. The choice of form of economic evaluation is justified in relation to the questions addressed.	¨	¨	√	
Data collection				
8. The source(s) of effectiveness estimates used are stated.	√	¨	¨	
9. Details of the design and results of effectiveness study are given (if based on a single study).	¨	¨	¨	√
10. Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies).	¨	√	¨	¨
11. The primary outcome measure(s) for the economic evaluation are clearly stated.	√	¨	¨	
12. Methods to value benefits are stated.	√	¨	¨	¨
13. Details of the subjects from whom valuations were obtained were given.	√	¨	¨	¨
14. Productivity changes (if included) are reported separately.	¨	¨	¨	√
15. The relevance of productivity changes to the study question is discussed.	¨	¨	¨	√
16. Quantities of resource use are reported separately from their unit costs.	¨	√	¨	

17.Methods for the estimation of quantities and unit costs are described.	¨	√	¨	
18.Currency and price data are recorded.	¨	√	¨	
19.Details of currency of price adjustments for inflation or currency conversion are given.	√	¨	¨	
20.Details of any model used are given.	√	¨	¨	¨
21.The choice of model used and the key parameters on which it is based are justified.	√	¨	¨	¨
Analysis and interpretation of results				
22.Time horizon of costs and benefits is stated.	√	¨	¨	¨
23.The discount rate(s) is stated.	√	¨	¨	¨
24.The choice of discount rate(s) is justified.	¨	¨	¨	√
25.An explanation is given if costs and benefits are not discounted.	√	¨	¨	¨
26.Details of statistical tests and confidence intervals are given for stochastic data.	¨	¨	√	¨
27.The approach to sensitivity analysis is given.	√	¨	¨	¨
28.The choice of variables for sensitivity analysis is justified.	√	¨	¨	¨
29.The ranges over which the variables are varied are justified.	√	¨	¨	¨
30.Relevant alternatives are compared.	√	¨	¨	¨
31.Incremental analysis is reported.	√	¨	¨	¨
32.Major outcomes are presented in a disaggregated as well as aggregated form.	¨	√	¨	
33.The answer to the study question is given.	√	¨	¨	
34.Conclusions follow from the data reported.	√	¨	¨	
35.Conclusions are accompanied by the appropriate caveats.	√	¨	¨	

Appendix 5 Data extraction results

Appendix 5.1 Study characteristics of the 15 effectiveness studies with two-step tests

(1) Data extraction result for Arora *et al.* (2013)

1. Study details	
Study ID: 01 **First author surname:** Arora **Year of publication:** 2013 **Country and city:** Thailand, Lampang **Study design:** Cross-sectional study **Study setting:** the antenatal care clinic of Lampang Regional Hospital **Number of centres:** 1 **Inclusion and exclusion criteria**: All pregnant women of appropriate gestational age for screening during the study period were included. **Time of study:** Between 4 January to 30 September 2010 **Funding:** Information not available	
2. Aim of the study	
This study aims to assess the GDM prevalence by using universal screening, and to show the diagnostic value of selective screening at Lampang Hospital.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	613
Number of participants consented/included	613
Number of participants in the end after those missed/excluded/lost to	593 (18 were exclude either because they were lost to follow-up or could not

follow-up at a later stage	complete the screening protocol)	
Characteristics of participants		
Mean age (years)	N/A (29.7% ≥30 years old)	
Mean BMI (kg/m^2)	N/A (21.9% ≥25)	
Ethnicity	95.1% Thai/ Chinese Thai, 2.0% Hill tribe, 2.9% others	
Participants with a family history of diabetes among first-degree relatives	13.3%	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	8	
Maternal age	Yes (≥30)	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes (≥4000g)	
History of adverse obstetric outcome	Yes (history of dead fetus in utero, fetal anomaly)	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Glucosuria Hypertension	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	593	593 (same cohort)
Selection		
Number of high risk women		313
Number of low risk women		280
Screening		
Number of women undertook screening	593	313

Time of screening	Women at risk: at first prenatal visits, and again at 24 weeks or more of gestation if the test was negative Other women: 24 weeks or more of gestation	Women at risk: at first prenatal visits, and again at 24 weeks or more of gestation if the test was negative
Screening test	50g GCT: ≥ 140 mg/dl at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 95 mg/dl, 1-hour 180 mg/dl, 2-hour 155 mg/dl and 3-hour 140 mg/dl)	Same as universal group
Diagnosis criteria reference	ACOG criteria	Same as universal group
Number of women diagnosed positive-GDM women (%) (GDM prevalence)	55/593 (9.3%)	43/593 (7.3%)
Sensitivity	100% (reference)	78.2%=43/55
Specificity	0% (reference)	49.8%=(280-12)/(593-55)
5.3 Other information		
None.		
6. Authors conclusion		
21.8% GDM cases might be missed if using risk-based screening, this might not be acceptable. The risk-based screening policy adopted throughout the country needs to be re-evaluated.		
7. Reviewer's conclusion		
Though selective screening based on risk factors could exempt 49.8% non-GDM women from screening, 21.8% GDM cases would be missed by the selective approach. The conclusion of 'risk-based screening might not be acceptable and needs to be re-evaluated' is an accurate reflection of the results presented.		

(2) Data extraction result for Caliskan *et al.* (2004)

1. Study details	
Study ID: 02 **First author surname:** Caliskan **Year of publication:** 2004 **Country and city:** Turkey, Ankara **Study design:** Retrospective cohort study **Study setting:** Antenatal policlinics of the Social Security Council Maternity and Women's Health Teaching Hospital **Number of centres:** 1 **Inclusion and exclusion criteria**: Pregnant women having singleton pregnancies between 24 and 28 weeks of gestation without a previous diagnosis of diabetes mellitus **Time of study:** Between May to July 2000 **Funding:** Information not available	
2. Aim of the study	
This study aims to test whether a prospective application of a risk factor scoring can be an alternative screening strategy for diagnosing GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	425
Number of participants consented/included	425
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	422 (3 women could not tolerate a 100g OGTT, they all had a 50g GCT<7.2mmol/l)
Characteristics of participants	
Mean age (years)	24.9 ±4.9 (16-40) (48% ≥ 25 years old)
Mean BMI (kg/m^2)	23.9 ±3.5 (16.9-39.1) (31.55% ≥ 25)

Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	24.6%	
4. Selection criteria for high-risk women		
Reference of criteria	From a retrospective study they performed (unpublished data)	
Type of criteria	Risk factor ≥1 or ≥2	
Number of risk factors used	5	
Maternal age	Yes (≥25)	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	None	
A prior macrosomic infant	Yes (>4000g)	
History of adverse obstetric outcome	Yes (recurrent spontaneous abortions, fetal anomaly despite a normal karyotype and prior unexplained *in utero* fetal death at a gestational age ≥20 weeks)	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	None	
5. Comparison of universal versus selective screening		
5.1 Selection criteria: risk factor ≥ 1		
	Universal screening	**Selective screening**
Number of participants	422	422 (same cohort)
Selection		
Number of high risk women		297
Number of low risk women		125
Screening		
Number of women undertook screening	422	297
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥ 7.2 mmol/l (130 mg/dl) at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A

Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	1 week after screening	Same as universal group
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 5.8 mmol/l (105 mg/dl), 1-hour 10.6 mmol/l (190 mg/dl), 2-hour 9.2 mmol/l (165 mg/dl) and 3-hour 8.1 mmol/l (145 mg/dl)	Same as universal group
Diagnosis criteria reference	National Diabetes Data Group	Same as universal group
Number of women diagnosed positive-GDM women (%)	14 (3.3%)	14 (3.3%)
Sensitivity	100% (reference)	100%=14/14
Specificity	0% (reference)	30.6%=(125-0)/(422-14)
5.2 Selection criteria: risk factor ≥ 2		
	Universal screening	**Selective screening**
Number of participants	422	422 (same cohort)
Selection		
Number of high risk women		157
Number of low risk women		265
Screening		
Number of women undertook screening	422	422
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥ 7.2 mmol/l (130 mg/dl) at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	1 week after screening	Same as universal group

Diagnosis test	100g OGTT (two or more values above thresholds): fasting 5.8 mmol/l (105 mg/dl), 1-hour 10.6 mmol/l (190 mg/dl), 2-hour 9.2 mmol/l (165 mg/dl) and 3-hour 8.1 mmol/l (145 mg/dl)	Same as universal group
Diagnosis criteria reference	National Diabetes Data Group	Same as universal group
Number of women diagnosed positive-GDM women (%) (GDM prevalence)	14/422 (3.3%)	12/422 (2.8%)
Sensitivity	100% (reference)	85.7%=12/14
Specificity	0% (reference)	64.5%=(265-2)/(422-14)
5.3 Other information		
Logistic regression analysis revealed that an increase in the score by one caused a three times increase in GDM risk (OR=3, CI=1.9–5).		
6. Authors conclusion		
A well integrated, population-based scoring will decrease the number of unnecessary testing but still diagnose 85–100% of GDM cases.		
7. Reviewer's conclusion		
This is a well-designed and clearly reported study. Using the risk scores ≥1 or ≥ 2, the sensitivity ranged from 100% to 85.7%, while the specificity ranged from 30.6% to 64.5%. It is rationale for the authors to conclude a well integrated risk scoring system based on the evidence of the population could reach a satisfying result. However, it is worth noticing that the selective criteria of risk scoring system are actually risk factors (women presenting ≥1 or ≥ 2 risk factors were deemed as high risk women).		

(3) Data extraction result for Coustan *et al.* (1989)

1. Study details
Study ID: 05 **First author surname:** Coustan **Year of publication:** 1989

Country and city: USA, Rhode Island

Study design: Prospective cohort study

Study setting: Four large obstetric practices in the region, and the obstetric clinic service

Number of centres: 5

Inclusion and exclusion criteria: Pregnant women (details not available)

Time of study: Between July 1 1984 to December 30 1986

Funding: Supported in part by the Centres for Disease Control (Atlanta, Georgia), the Rhode Island Department of Health, and the National Institutes of Health

2. Aim of the study	
This study aims to evaluate the sensitivity and cost-effectiveness of various screening schemes.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	6214
Number of participants consented/included	6106
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	5379 (108 women's questionnaires of risk factors were not available)
Characteristics of participants	
Mean age (years)	N/A (27%≥30 years old)
Mean BMI (kg/m^2)	N/A
Ethnicity	87.9% white, 8.1% black, 1.2% Asian, 0.5% Indian, 0.1% Chinese, 2.0% unkown
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	The American College of Obstetricians and Gynecologists (ACOG)
Type of criteria	Risk factor ≥1

Number of risk factors used	6	
Maternal age	Yes (≥30)	
Obesity	Yes (weight ≥85th percentile for height)	
Family history of diabetes among first-degree relatives	Yes (among first- or second-degree relatives)	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes (≥ 9 lb)	
History of adverse obstetric outcome	Yes (previous stillborn, neonatal death, or preterm delivery)	
Member of an ethnic or racial group with a high prevalence of GDM	None	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	6106	6106 (same cohort)
Selection		
Number of high risk women		2680
Number of low risk women		3426
Screening		
Number of women undertook screening	6106	2680
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥130 mg/dl at 1 hour	Same as universal group
Screening criteria reference	N/A	Same as universal group
Number of women screened positive (%)	N/A	Same as universal group
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	N/A	Same as universal group

Diagnosis test	100g OGTT	Same as universal group
Diagnosis criteria reference	The ACOG and the National Diabetes Data Group	Same as universal group
Number of women diagnosed positive-GDM women (%)	125/6106 (2.0%)	81/6106 (1.1%)
Sensitivity	100% (reference)	64.8%=81/125
Specificity	0% (reference)	56.5%=(3426-44)/(6106-125)
5.3 Other information		
None.		
6. Authors conclusion		
If the ACOG recommendation of selective screening was used, 35% GDM women would be missed, with little cost savings.		
7. Reviewer's conclusion		
The sensitivity of 65% (35% GDM would be missed) was estimated from the general population of pregnant women, not calculated from the GDM population. The figure of 108 (1.8%) women whose risk factors were not available was very small to affect the accuracy of study results of specificity and sensitivity.		

(4) Data extraction result for Danilenko-Dixon *et al.* (1999)

1. Study details
Study ID: 06 **First author surname:** Danilenko-Dixon **Year of publication:** 1999 **Country and city:** USA, Rochester **Study design:** Retrospective cohort study

Study setting: Mayo Clinic in Rochester

Number of centres: 1

Inclusion and exclusion criteria: All women receiving obstetric care at the Mayo Clinic (Rochester, Minnesota)

Time of study: Between 1 January 1986 to 31 December 1997

Funding: Information not available

2. Aim of the study	
This study aims to evaluate the impact of the 1997 American Diabetes Association selective screening guidelines for GDM applied to a universally screened population.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	18834
Number of participants consented/included	18504
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	18174 (330 women had one or more of the risk factors data according to the ADA low-risk criteria missed)
Characteristics of participants	
Mean age (years)	28.8±5.3 (82.2%≥ 30 years old; 51.2%≥ 25 years old)
Mean BMI (kg/m^2)	24.3±5.1 (33%≥ 25; 22%≥ 27)
Ethnicity	92.6% white, 0.7% African American, 4.9% Asian, 1.7% Hispanic, 0.1% Other
Participants with a family history of diabetes among first-degree relatives	7.7%
4. Selection criteria for high-risk women	
Reference of criteria	1997 American Diabetes Association GDM screening guidelines (same with the Fourth International Workshop-Conference on GDM)
Type of criteria	Risk factor ≥1
Number of risk factors used	4
Maternal age	Yes (≥25)
Obesity	Yes (BMI ≥27)

Family history of diabetes among first-degree relatives	Yes (among first degree relatives)	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	Yes	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	18504	18504 (same cohort)
Selection		
Number of high risk women		16672
Number of low risk women		1832
Screening		
Number of women undertook screening	18504	16672
Time of screening	24-30 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥140 mg/dl at 1 hour	Same as universal group
Screening criteria reference	N/A	Same as universal group
Number of women screened positive (%)	3683 (20%)	N/A
Diagnosis		
Number of women undertook diagnosis	3683	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT(two or more values above thresholds): fasting 5.9 mmol/l (105 mg/dl), 1-hour 10.6 mmol/l (190 mg/dl), 2-hour 9.2 mmol/l (165 mg/dl) and 3-hour 8.1 mmol/l (145 mg/dl)	Same as universal group

Diagnosis criteria reference	National Diabetes Data Group criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	564/18504 (3.0%)	547/18504 (3.0%)
Sensitivity	100% (reference)	97%=547/564
Specificity	0% (reference)	10.1%=(1832-17)/(18504-564)
5.3 Other information		
Screening only women≥ 25 years old would have detected 90.4% of GDM cases, whereas the addition of the remaining 3 screening criteria combined would have detected only an additional 6.6% of cases. Screening only women≥ 25 years old would have exempted 17.8% women. Screening only women≥ 30 years old alone would have detected 65.8% of GDM cases, and exempted 48.8% women.		
6. Authors conclusion		
The 1997 American Diabetes Association selective screening guidelines would miss relatively small proportion (3%) of GDM women in our population. However, implementation of these guidelines would decrease the number of screens by only 10% while adding significant complexity to the screening process.		
7. Reviewer's conclusion		
This is a well reported study. The study additionally analysed the significance of each of the four risk factors for predicting GDM, and found age is the most significant risk factor for GDM in their population. The cut-off value of 25 years old is more appropriate than 30 years old as age criteria in their population.		

(5) Data extraction result for Di Cianni *et al.* (2003)

1. Study details
Study ID: 07 **First author surname:** Di Cianni **Year of publication:** 2003 **Country and city:** Italy, Pisa

Study design: Retrospective cohort study

Study setting: Diabetes Section of the Department of Endocrinology and Metabolism of the University of Pisa

Number of centres: 1

Inclusion and exclusion criteria: Pregnant women consecutively referred to the Diabetes Section of the Department of Endocrinology and Metabolism of the University of Pisa

Time of study: Between 1 June 1995 to 31 December 2001

Funding: Information not available

2. Aim of the study

This study aims to evaluate the GDM prevalence and the presence of risk factors for GDM, as well as to compare universal versus selective screening to validate the ADA's recommendations in their population.

3. Participants

Number of participants (One cohort of universal screening)	
Number of participants invited/considered	3950
Number of participants consented/included	3806
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	3662 (144 women with GCT positive didn't go for OGTT test)
Characteristics of participants	
Mean age (years)	31.1 ±4.7 (15-50)
Mean BMI (kg/m^2)	22.5 ±3.7 (15.6-48.8)
Ethnicity	Caucasian
Participants with a family history of diabetes among first-degree relatives	18.1%
4. Selection criteria for high-risk women	
Reference of criteria	2002 American Diabetes Association (ADA) GDM screening guidelines
Type of criteria	Risk factor ≥1
Number of risk factors used	3
Maternal age	Yes (≥25)

Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	3950	3950 (same cohort)
Selection		
Number of high risk women		3714
Number of low risk women		236
Screening		
Number of women undertook screening	3950	3714
Time of screening	24-28 weeks of gestation (the obstetrician could request the GCT at 14-18 weeks for women at risk, and if the GCT is negative, it will be further performed at 24-28 weeks)	Same as universal group
Screening test	50g GCT: ≥7.8 mmol/l at 1 hour	Same as universal group
Screening criteria reference	Recommendations of the Regional Public Health Authority, and the Fourth International Workshop Conference on GDM	Same as universal group
Number of women screened positive (%)	1389 (35.2%)	1355
Diagnosis		
Number of women undertook diagnosis	1221 (Among the 1389 GCT+ women, 24 were diagnosed as GDM directly, 144 dropped out)	1355

Time of diagnosis	One week after 50g GCT	Same as universal group
Diagnosis test	50g GCT ≥11 mmol/l allowed a direct diagnosis of GDM 100g OGTT(two or more values above thresholds): fasting 5.9 mmol/l (105 mg/dl), 1-hour 10.6 mmol/l (190 mg/dl), 2-hour 9.2 mmol/l (165 mg/dl) and 3-hour 8.1 mmol/l (145 mg/dl)	Same as universal group
Diagnosis criteria reference	Carpenter and Coustan's criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	308/3806 (8.1%)	303/3806(8.0%)
Sensitivity	100% (reference)	98.4%=303/308
Specificity	0% (reference)	6.6%=(236-5)/(3806-308)
5.3 Other information		
None.		
6. Authors conclusion		
At present, universal screening remains a good instrument to identify women with GDM at least in Italian with a large percentage of women at medium/high risk for GDM.		
7. Reviewer's conclusion		
This is a well reported study. The study additionally performed OGTT among 391 randomly selected GCT negative women, and diagnosed 25 more GDM women. However, to keep GDM prevalence calculation consistent, these 25 GDM women were excluded for the calculation of GDM prevalence (Only the 308 GDM women who were diagnosed under the normal procedure of GCT positive followed by OGTT were calculated).		

(6) Data extraction result for Hadaegh *et al.* (2005)

1. Study details	
Study ID: 09 **First author surname:** Hadaegh **Year of publication:** 2005 **Country and city:** Iran, Bandar Abbas **Study design:** Prospective cohort study **Study setting:** Obstetrics clinics in various parts of Bandar Abbas city in southern Iran **Number of centres:** many **Inclusion and exclusion criteria**: Pregnant women consecutively referred to the obstetrics clinics were included. Women with any of the following characteristics were excluded from the study: (1) previous history of diabetes; (2) use of drugs that affect glucose metabolism, such as corticosteroids; (3) diagnosis of chronic liver disease, endocrine disorders (such as hyperthyroidism), or connective tissue disorders; and (4) presence of major medical conditions, such as persistent hypertension. **Time of study:** Between March 2002 to March 2004 **Funding:** Information not available	
2. Aim of the study	
This study aims to estimate and report the prevalence of GDM in pregnant women of Bandar Abbas, a city in southern Iran.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	800
Number of participants consented/included	700
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	600 (100 women who had GCT positive results did not undergo the 3-hour OGTT mainly because of personal reasons. No significant differences were found in age, BMI, or blood pressure between those pregnant women who

	underwent the OGTT and those who did not)	
Characteristics of participants		
Mean age (years)	24.9±5.3	
Mean BMI (kg/m^2)	BMI≥25 kg: 48.9% in normal women and 67.7% in GDM women	
Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	9.2% in normal women and 12.9% in GDM women	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	3	
Maternal age	Yes (≥25)	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	800	800 (same cohort)
Selection		
Number of high risk women		552
Number of low risk women		248
Screening		
Number of women undertook screening	800	552
Time of screening	24-28 weeks of gestation	Same as universal group

Screening test	50g GCT: ≥130 mg/dl at 1 hour	Same as universal group
Screening criteria reference	N/A	Same as universal group
Number of women screened positive (%)	221/700 (31.6%)	157
Diagnosis		
Number of women undertook diagnosis	221	157
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT(two or more values above thresholds): fasting 95 mg/dl, 1-hour 180 mg/dl, 2-hour 155 mg/dl and 3-hour 140 mg/dl	Same as universal group
Diagnosis criteria reference	Carpenter and Coustan's criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	62/700 (8.9%)	55/700 (7.9%)
Sensitivity	100% (reference)	88.7%=55/62
Specificity	0% (reference)	32.7%=(248-7)/(800-62)
5.3 Other information		
Of the 800 participants, 100 women who had a positive GCT results did not undergo OGTT, which could lead to underestimation of GDM prevalence. To avoid this, the predicted prevalence of GDM in these women was calculated, and the overall GDM prevalence was estimated to be 11.4% (91 of 800) under the Carpenter and Coustan's criteria of GDM diagnosis.		
6. Authors conclusion		
Screening for GDM in all pregnant women in Bandar Abbas seems necessary, regardless of the presence of risk factors for GDM.		
7. Reviewer's conclusion		
This is a well reported study. The study analysed the GDM prevalence using the National Diabetes Data Group and the Carpenter and Coustan criteria, respectively. However, as the sensitivity was calculated using the Carpenter and Coustan criteria, only data under the Carpenter and Coustan criteria were extracted.		

(7) Data extraction result for Helton *et al.* (1997)

1. Study details	
Study ID: 10 **First author surname:** Helton **Year of publication:** 1997 **Country and city:** USA, Chapel Hill **Study design:** Retrospective cohort study **Study setting:** A university-based family practice centre **Number of centres:** 1 **Inclusion and exclusion criteria**: Pregnant women who received prenatal care and gave birth at the centre during the study period, and had no personal history of diabetes. **Time of study:** Between January 1988 to December 1993 **Funding:** The study was supported by a University of North Carolina Department of Family Medicine Faculty Research Development Grant.	
2. Aim of the study	
This study aims to reassess the value of universal screening of all pregnant patients with a 50g GCT.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	595
Number of participants consented/included	526
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	457 (Among these excluded 69 women, 51 were either not screened by the GCT test or had no record of a screening result, and 18 were those who had a GCT positive results but did not undergo OGTT)
Characteristics of participants	
Mean age (years)	27 (15-46)

Mean BMI (kg/m^2)	N/A	
Ethnicity	60% non-Hispanic white, 35% black, 5% other races	
Participants with a family history of diabetes among first-degree relatives	N/A	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	6	
Maternal age	Yes (≥35)	
Obesity	Yes (weight ≥200lb)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes	
History of adverse obstetric outcome	Yes	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	595	595 (same cohort)
Selection		
Number of high risk women		189
Number of low risk women		406
Screening		
Number of women undertook screening	544	189
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥140 mg/dl at 1 hour	Same as universal group
Screening criteria reference	The American College of Obstetricians and Gynecologists (ACOG) criteria	Same as universal group

Number of women screened positive (%)	76/544 (12.8%)	N/A
Diagnosis		
Number of women undertook diagnosis	58 (18 GCT positive women did not undergo OGTT)	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT(two or more values above thresholds): fasting 105 mg/dl, 1-hour 190 mg/dl, 2-hour 165 mg/dl and 3-hour 145 mg/dl	Same as universal group
Diagnosis criteria reference	National Diabetes Data Group (NDDG) criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	13/595 (2.2%)	9/595 (1.5%)
Sensitivity	100% (reference)	69.2%=9/13
Specificity	0% (reference)	69.1%=(406-4)/(595-13)
5.3 Other information		
None.		
6. Authors conclusion		
Screening only those women at risk for GDM is a reasonable approach to identify the disease in a low-risk population.		
7. Reviewer's conclusion		
The specificity of 69.1% which could exempt 69.1% non-GDM women from screening was outstanding. Although 30.8% GDM women would be missed by selective screening, however, the absolute value of missing 4 GDM women of 595 pregnant women is small in this low risk population (GDM prevalence was 2.2%). It is considered rational that the author concluded selective screening was a preferable alternative to universal screening in low risk population.		

(8) Data extraction result for Jimenez-Moleon *et al.* (2002)

1. Study details
Study ID: 12 **First author surname:** Jimenez-Moleon **Year of publication:** 2002 **Country and city:** Spain, Granada **Study design:** Retrospective cohort study **Study setting:** San Cecilio University Hospital of Granada (SCUH) **Number of centres:** 1 **Inclusion and exclusion criteria**: Criteria for inclusion in the study were: (a) having regular residence and medical attention within the area of referral of the SCUH, (b) having a singleton pregnancy, (c) making a first visit to the doctor before week 28 of gestation and (d) gestational age at delivery ≥28 weeks. The exclusion criteria were: primary diabetes (type 1 or type 2) or carbohydrate intolerance diagnosed before gestation, a pregnancy not under medical control and pregnancies and deliveries involving a high obstetric risk which, under other circumstances, would not have been attended to in the SCUH. **Time of study:** Between 1 January and 31 December 1995 **Funding:** The study was supported was supported in part by the Fondo de Investigaciones Sanitarias (Fund for Health Research in Spain).
2. Aim of the study
This study aims to determine the prevalence of GDM and its variations depending on the presence of risk factors, and to evaluate how the GDM screening strategies applied might modify the observed prevalence in the population.
3. Participants

Number of participants (One cohort of universal screening)	
Number of participants invited/considered	2574
Number of participants consented/included	1962 (76.2%)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1350 (for the other 612 women, screening was definitely not done in 425 and was not known in 187 women)

Characteristics of participants		
Mean age (years)	80.3%≥ 25 years, 41.8%≥ 30 years	
Mean BMI (kg/m^2)	12.3%≥27	
Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	14.8%	
4. Selection criteria for high-risk women		
Reference of criteria	The American Diabetes Association (ADA) criteria and the American College of Obstetrics and Gynecologists (ACOG) criteria, respectively (The only different between the two criteria is the age threshold).	
Type of criteria	Risk factor ≥1	
Number of risk factors used	10	
Maternal age	Yes (≥25 for ADA criteria and ≥30 for ACOG criteria)	
Obesity	Yes (BMI ≥27)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes	
History of adverse obstetric outcome	Yes (two or more miscarriages, perinatal mortality or congenital malformations)	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Chronic hypertension	
	Polyhydramnios	
	Hpertension induced by the pregnancy	
	Suspected large fetus for gestational age	
5.1 Comparison of universal versus selective screening (under ACOG criteria for selecting high risk women)		
	Universal screening	**Selective screening**
Number of participants	2574	2574 (same cohort)
Selection		
Number of high risk women		1436

Number of low risk women		1138
Screening		
Number of women undertook screening	1962	1162
Time of screening	24-28 weeks of gestation	For women with risk factors, performed at the first medical visit during pregnancy, and if negative then repeated after week 24 of gestation
Screening test	50g GCT: ≥7.8 mmol/l at 1 hour	Same as universal group
Screening criteria reference	N/A	Same as universal group
Number of women screened positive (%)	294/1962 (15.0%)	N/A
Diagnosis		
Number of women undertook diagnosis	259 (80% of 294 women with GCT+ underwent OGTT)	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 5.8 mmol/l, 1-hour 10.6mmol/l, 2-hour 9.2mmol/l and 3-hour 8.1mmol/l	Same as universal group
Diagnosis criteria reference	The National Diabetes Data Group critieria	Same as universal group
Number of women diagnosed positive-GDM women (%)	65/2574 (2.5%)	58/2574 (2.3%)
Sensitivity	100% (reference)	89.2%=58/65
Specificity	0% (reference)	45.1%=(1138-7)/(2574-65)
5.2 Comparison of universal versus selective screening (under ADA criteria for selecting high risk women)		
	Universal screening	**Selective screening**

Number of participants	2574	2574 (same cohort)
Selection		
Number of high risk women		2174
Number of low risk women		400
Screening		
Number of women undertook screening	1962	1704
Time of screening	24-28 weeks of gestation	For women with risk factors, performed at the first medical visit during pregnancy, and if negative then repeated after week 24 of gestation
Screening test	50g GCT: ≥7.8 mmol/l at 1 hour	Same as universal group
Screening criteria reference	N/A	Same as universal group
Number of women screened positive (%)	294/1962 (15.0%)	N/A
Diagnosis		
Number of women undertook diagnosis	259 (80% of 294 women with GCT+ underwent OGTT)	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT(two or more values above thresholds): fasting 5.8 mmol/l, 1-hour 10.6mmol/l, 2-hour 9.2mmol/l and 3-hour 8.1mmol/l	Same as universal group
Diagnosis criteria reference	The National Diabetes Data Group critieria	Same as universal group
Number of women diagnosed positive-GDM women (%)	65/2574 (2.5%)	63/2574 (2.4%)
Sensitivity	100% (reference)	96.9%=63/65

Specificity	0% (reference)	15.9%=(400-2)/(2574-65)
5.3 Other information		
None.		
6. Authors conclusion		
Selective GDM screening under ADA criteria did not have any apparent benefits. Selective screening is desirable only when fairly restrictive criteria are applied in defining the gravidae at risk and, therefore, a significant proportion of the population is exempt from screening.		
7. Reviewer's conclusion		
The study evaluated the specificity and selectivity of the selective screening using the ADA criteria (age cut-off 25 years or over) and the ACOG criteria (age cut-off 30 years or older) respectively. The only difference of the two selection criteria is the age cut-off. As a result, the specificity (45.1% and 15.9%, respectively) and the selectivity (89.2% and 96.9%, respectively) changed considerably.		

(9) Data extraction result for Lavin (1985)

1. Study details
Study ID: 13 **First author surname:** Lavin **Year of publication:** 1985 **Country and city:** USA, Akron (Ohio) **Study design:** Prospective cohort study **Study setting:** the prenatal clinic of the University of Cincinnati Medical Center **Number of centres:** 1 **Inclusion and exclusion criteria**: Women who were not known to be diabetic before the onset of pregnancy during a 25-month period **Time of study:** Information not available

Funding: Information not available	
2. Aim of the study	
This study aims to evaluate whether screening only the high-risk women is appropriate and to determine if universal screening for abnormal carbohydrate metabolism in pregnancy could be feasibly implemented in a busy, working prenatal clinic.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	2077
Number of participants consented/included	2077
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	2077
Characteristics of participants	
Mean age (years)	N/A
Mean BMI (kg/m^2)	N/A
Ethnicity	N/A
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	N/A
Type of criteria	Risk factor ≥1
Number of risk factors used	9
Maternal age	No
Obesity	Yes
Family history of diabetes among first-degree relatives	Yes (A family member with diabetes)
Personal history of GDM	Yes
A prior macrosomic infant	Yes (>4000g)
History of adverse obstetric outcome	Yes (previous delivery of an unexplained stillborn infant, previous delivery of an infant with significant congenital malformations, a history of recurrent abortion)

Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Monilial vaginitis	
	Glucosuria	
	Polyhydramnios	
	An infant suspected of being large for gestational age	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	2077	2077 (same cohort)
Selection		
Number of high risk women		959
Number of low risk women		1118
Screening		
Number of women undertook screening	2077	959
Time of screening	28-32 weeks of gestation for non-risk women; at the initial visit and again at 28 weeks of gestation for the at-risk women	At the initial visit and again at 28 weeks of gestation
Screening test	50g GCT: ≥150 mg/dl at 1 hour	Same as universal group
Screening criteria reference	O'Sullivan *et al.*, and the American Diabetes Association Workshop-Conference on Gestational Diabetes	Same as universal group
Number of women screened positive (%)	137 (6.6%)	69 (7.2%)
Diagnosis		
Number of women undertook diagnosis	137	69
Time of diagnosis	N/A	N/A
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 105 mg/dl, 1-hour 190 mg/dl,	Same as universal group

	2-hour 165 mg/dl and 3-hour 145 mg/dl)	
Diagnosis criteria reference	O'Sullivan *et al.*, and the American Diabetes Association Workshop-Conference on Gestational Diabetes	Same as universal group
Number of women diagnosed positive-GDM women (%)	30/2077 (1.4%)	14/2077 (0.7%)
Sensitivity	100% (reference)	46.7%=14/30
Specificity	0% (reference)	53.8%=(1118-16)/(2077-30)
5.3 Other information		
The results of the study have been reported previously (Lavin *et al*., 1982), and were summarised in this article. The previous article (Lavin *et al*., 1982) was not accessible, thus could not identify any possible information about the time of study and the characteristics of participants, which might be available in that article.		
6. Authors conclusion		
The study results reemphasize the inadequacy of screening only those patients with traditional risk factors for GDM and demonstrate the feasibility of implementing a program of universal glucose screening among a large obstetric population.		
7. Reviewer's conclusion		
The author divided the cohort into risk women and non-risk women. However, it was still a one-cohort study. If screening high-risk women only, it could exempt 53.8% non-GDM women from screening, but would miss more than half (53.3%) of the GDM cases. It is rationale for the author to conclude that screening only high-risk women for GDM in their population was inadequate.		

(10) Data extraction result for Naylor *et al.* (1997)

1. Study details

Study ID: 17

First author surname: Naylor

Year of publication: 1997

Country and city: Canada, Toronto

Study design: Retrospective cohort study

Study setting: Three teaching hospitals in Toronto

Number of centres: 3

Inclusion and exclusion criteria: All consenting pregnant women without known diabetes mellitus, 24 years of age or older, presenting before 24 weeks' gestation, had singleton deliveries, had both the screening and diagnosis tests, and had complete data during the study period were included. Women who delivered before 28 weeks' gestational were excluded.

Time of study: Between September 1989 to March 1992

Funding: The study was supported by an operating grant (02650) from the Ontario Ministry of Health

2. Aim of the study

This study aims to test the hypothesis that the efficiency of screening could be enhanced by considering women's risks of GDM on the basis of their clinical characteristics.

3. Participants

Number of participants (One cohort of universal screening)	
Number of participants invited/considered	3152
Number of participants consented/included	3131 (99.3%) (divided into two groups: 1560 as derivation group and 1571 was validation group)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	3110 (The other 21 either did not have singleton deliveries or data were insufficiently complete for analysis)

Characteristics of participants		
Mean age (years)	51.3% ≥ 30 (for the derivation group)	
Mean BMI (kg/m^2)	47.2% ≥ 22 (for the derivation group)	
Ethnicity	81.5% White, 5.3% Black, 9.0% Asian, 4.3% others	
Participants with a family history of diabetes among first-degree relatives	14.5%	
4. Selection criteria for high-risk women		
Reference of criteria	A risk score system derived from their derivation group (1560 women) using the rounded odds ratios of the risk factors	
Type of criteria	A risk score system (women with scores of 0 to 1 were not screened but the reminder were all screened)	
Number of risk factors used	3	
Maternal age	Yes (scored 0 if age ≤30; scored 1 if age within 31-34; scored 2 if age ≥35)	
Obesity	Yes (scored 0 if BMI ≤22; scored 2 if BMI within 22.1-25.0; scored 3 if BMI ≥25.1)	
Family history of diabetes among first-degree relatives	No	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	Yes (scored 0 if black; scored 5 if Asian; scored 2 if others)	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1560 (derivation group)	1560 (same cohort)
Selection		
Number of high risk women		1016
Number of low risk women		544

Screening		
Number of women undertook screening	1560	1016
Time of screening	26 weeks' gestation (±1 week)	Same as universal group
Screening test	50g GCT: ≥ 140 mg/dl for women scored 2-3; ≥ 128 mg/dl for women scored ＞3 at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	28 weeks' gestation (±1 week)	Same as universal group
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 5.8 mmol/l (105 mg/dl), 1-hour 10.5 mmol/l (190 mg/dl), 2-hour 9.2 mmol/l (165 mg/dl) and 3-hour 8.0 mmol/l (145 mg/dl)	Same as universal group
Diagnosis criteria reference	N/A	Same as universal group
Number of women diagnosed positive-GDM women (%)	32/1560 (2.1%)	29/1560 (1.9%)
Sensitivity	100% (reference)	90.6%=29/32
Specificity	0% (reference)	35.4%=(544-3)/(1560-32)
5.3 Other information		
For the derivation group, apart from the 50g GCT thresholds extracted as above, the author also examined the efficacy of selective screening using different 50g GCT thresholds, namely, test threshold of 140 mg/dl for scores 2–3, 130 mg/dl for scores ＞3). The GDM diagnosed would be 27		
Funding: Information not available		

6. Authors conclusion
Consideration of women's clinical characteristics allows efficient selective screening for GDM.
7. Reviewer's conclusion
The specificity was not directly reported from the study, but calculated by the reviewer from the data of the derivation group (32 GDM cases detected in the universal screening group of usual care, and 29 cases detected in the selective screening group using 140 mg/dl threshold for women scored 2-3 and 128mg/dl threshold for women scored >3). Because of this reason, the reviewer only extracted the data from the derivation group (1560 women) where specificity could be calculated, data from the validation group (1571) were not extracted. This study was the first study which used a risk score system rather than presence/absence of risk factors for selective screening of GDM. This study also showed that the sensitivity of selective screening could change according to the different thresholds of the 50g GCT screening test.

(11) Data extraction result for Sacks *et al.* (1987)

1. Study details
Study ID: 18 **First author surname:** Sacks **Year of publication:** 1987 **Country and city:** USA, California **Study design:** Prospective cohort study **Study setting:** Kaiser Foundation Hospital, Bellflower, California **Number of centres:** 1 **Inclusion and exclusion criteria**: All pregnant women who delivered at the study period, had at least one glucose screening test during pregnancy, and had the presence or absence of risk factors ascertained. **Time of study:** Between July 1, 1984 and June 30, 1985
2. Aim of the study

This study aims to test the hypothesis that the efficiency of screening could be enhanced by considering women's risks of GDM on the basis of their clinical characteristics.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	4519
Number of participants consented/included	4116 (91.1%)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	3713 (For the other 403 women, 370 didn't have glucose screening test, 33 didn't have the profile of risk factors)
Characteristics of participants	
Mean age (years)	26.4
Mean BMI (kg/m^2)	N/A
Ethnicity	N/A
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	N/A
Type of criteria	Risk factor ≥1
Number of risk factors used	5 (previous pregnancies complicated by congenital malformations, stillbirth, or neonatal death were group as 'history of adverse obstetric outcome')
Maternal age	Yes (≥25)
Obesity	Yes (pre-pregnancy weight ≥150 pounds)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	No
A prior macrosomic infant	Yes
History of adverse obstetric outcome	Yes (including previous pregnancies complicated by congenital malformations, stillbirth, or neonatal death)

Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	4116	4116 (same cohort)
Selection		
Number of high risk women		3180
Number of low risk women		936
Screening		
Number of women undertook screening	4116	4116 (same cohort)
Time of screening	At or beyond 24 weeks of gestation	At the time of prenatal registration (initial clinic visit)
Screening test	50g GCT: ≥ 135 mg/dl	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	991	874
Diagnosis		
Number of women undertook diagnosis	N/A (11-21% women screened positive didn't go for OGTT)	N/A
Time of diagnosis	Approximately 18 days after the 50g GCT	Same as universal group
Diagnosis test	75g OGTT (If the fasting plasma glucose ≥ 120 mg/dL, the test was discontinued. If a subsequent fasting plasma glucose ≥105 mg/dL, the patient was considered a GDM)	Same as universal group

Diagnosis criteria reference	The Second International workshop-Conference on Gestational Diabetes, with one modification	Same as universal group
Number of women diagnosed positive-GDM women (%)	138/4116 (3.4%)	134/4116 (3.3%)
Sensitivity	100% (reference)	97.1%=134/138
Specificity	0% (reference)	23.4%=(936-4)/(4116-138)
5.3 Other information		
For pregnant women with risk factors, the GDM screening test was undertaken in early pregnancy (at the time of prenatal registration). Among patients whose early screening values were elevated and whose initial OGTT were normal, the odds of being classified ultimately as a GDM were 7.3 times that of patients whose initial screening tests were normal.		
6. Authors conclusion		
Selective screening based on risk factors may enhance detection of diabetes early in gestation, and might be considered in designing a cost-effective screening programme for GDM.		
7. Reviewer's conclusion		
Although the screening test for high risk pregnant women were conducted in early pregnancy (not during 24-28 weeks of gestation), the author had examined that screening test undertaken at this time point is effective as well (the figure of 7.3 times in the 'Other information'). In this study, it is rationale that the author concluded selective screening during early pregnancy is a good approach.		

(12) Data extraction result for Teh *et al.* (2011)

1. Study details	
Study ID: 19 **First author surname:** Teh **Year of publication:** 2011 **Country and city:** Australia, Victoria **Study design:** Retrospective cohort study **Study setting:** A tertiary referral hospital (Monash Medical Centre, Victoria) **Number of centres:** 1 **Inclusion and exclusion criteria**: Women of singleton pregnancies (without pre-existing diabetes mellitus) giving birth in 2007 from the Birthing Outcomes System. **Time of study:** Year 2007 **Funding:** The study was supported by a grant from the Novo Nordisk Regional Diabetes Scheme	
2. Aim of the study	
This study aims to review risk profiles of women with GDM and to evaluate international GDM screening recommendations.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	2880
Number of participants consented/included	2880
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	2880
Characteristics of participants	
Mean age (years)	33.3 ±5.2 (GDM women); 30.3 ±5.5 (Non-GDM women)
Mean BMI (kg/m^2)	28.2 ±8.3 (GDM women); 25.3 ±6.4 (Non-GDM women)

Ethnicity	56.9% were born overseas, 32.7% were from Asian countries
Participants with a family history of diabetes among first-degree relatives	39.8%
4. Selection criteria for high-risk women	
4.1 UK NICE selection criteria	
Reference of criteria	UK NICE
Type of criteria	Risk factor ≥1
Number of risk factors used	5
Maternal age	No
Obesity	Yes (BMI ≥30)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes (≥4500g)
History of adverse obstetric outcome	No
Member of an ethnic or racial group with a high prevalence of GDM	Yes (family origin with a high prevalence of diabetes including South Asian, black Caribbean, Middle Eastern)
Others	None
4.2 US ADA selection criteria	
Reference of criteria	US ADA
Type of criteria	Risk factor ≥1
Number of risk factors used	6
Maternal age	Yes (≥25)
Obesity	Yes (abnormal body weight)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	No

A prior macrosomic infant	No
History of adverse obstetric outcome	Yes
Member of an ethnic or racial group with a high prevalence of GDM	Yes (from a high-risk ethnic／racial group of diabetes, e.g. Hispanic American, Native American, Asian American, African American, Pacific Islander)
Others	History of abnormal glucose metabolism
4.3 Australian ADIPS selection criteria	
Reference of criteria	Australian ADIPS (universal screening is recommended, the selective screening criteria is for resource limited settings)
Type of criteria	N/A
Number of risk factors used	7
Maternal age	Yes (≥30)
Obesity	Yes
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes (Past history of GDM or glucose intolerance)
A prior macrosomic infant	No
History of adverse obstetric outcome	Yes
Member of an ethnic or racial group with a high prevalence of GDM	Yes (Belonging to high-risk ethnic group, e.g. Australian Indigenous, Polynesian, South Asian／Indian)
Others	Glycosuria

5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	2880	2880 (same cohort)

Selection		
Number of high risk women		1695/2880 (NICE criteria); 2340/2880 (ADA criteria); 2121/2880 (ADIPS criteria)
Number of low risk women		1185/2880 (NICE criteria); 540/2880 (ADA criteria); 759/2880 (ADIPS criteria)
Screening		
Number of women undertook screening	2880	2880 (same cohort)
Time of screening	26-28 weeks of gestation	26-28 weeks of gestation
Screening test	75g GCT: ≥ 8.0 mmol/l at 1 hour	Same as universal group
Screening criteria reference	The guidelines of the Australian Diabetes in Pregnancy Society	Same as universal group
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	N/A	N/A
Diagnosis test	75g OGTT (one or more values above thresholds): fasting 5.5 mmol/l, 2-hour 8.0 mmol/l	Same as universal group
Diagnosis criteria reference	The guidelines of the Australian Diabetes in Pregnancy Society	Same as universal group
Number of women diagnosed positive-GDM women (%)	250/2880 (8.7%)	232/2880 (8.1%) (NICE criteria); 250/2880 (8.7%) (ADA criteria); 247/2880 (8.6%) (ADIPS criteria)
Sensitivity	100% (reference)	92.7%=232/250 (NICE criteria); 100%=250/250 (ADA criteria); 98.6%=247/250 (ADIPS criteria)

Specificity	0% (reference)	44.4%=(1185-18)/(2880-250)=(NICE criteria); 2.1%(540-0)/(2880-250)(ADA criteria); 28.7%=(759-3)/(2880-250)=(ADIPS criteria)
5.3 Other information		
None.		
6. Authors conclusion		
Current selective screening guidelines have high sensitivity but low specificity and offer little over universal screening.		
7. Reviewer's conclusion		
The advantage of the study is that the author examined three different selective screening criteria, the UK NICE criteria, the US ADA criteria, and the Australian ADIPS criteria, respectively. All the three selective screening approach gave low specificity and high sensitivity (except the NICE criteria had lower sensitivity of 92.7% but still high). The three specificities reported in this table were based on the definition of specificity in the review, and were slightly different from the specificities reported by the author.		

(13) Data extraction result for Van Leeuwen *et al.* (2010)

1. Study details
Study ID: 20 **First author surname:** Van Leeuwen **Year of publication:** 2010 **Country and city:** Netherlands, Utrecht **Study design:** Prospective cohort study **Study setting:** University Medical Centre in Utrecht in the Netherlands **Number of centres:** 1

Inclusion and exclusion criteria: All women with a singleton pregnancy who reported for prenatal care during a period of 2 years were invited to participate in this study. Women with known pre-gestational diabetes mellitus and women who were first seen after 20 weeks of gestation were excluded from the study.

Time of study: N/A

Funding: The study was supported by a grant from Novo Nordisk, Alphen aan den Rijn, the Netherlands and by grant 917.46.346 in the VIDI-program of ZonMW, The Hague, the Netherlands.

2. Aim of the study	
This study aims to develop a clinical prediction rule that can help the clinician to identify women at high/low risk for GDM early in pregnancy in order to improve the efficiency of GDM screening.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	995
Number of participants consented/included	995
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	995
Characteristics of participants	
Mean age (years)	N/A
Mean BMI (kg/m^2)	N/A
Ethnicity	N/A
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	Their own clinical prediction model
Type of criteria	Risk factor ≥1
Number of risk factors used	4
Maternal age	No
Obesity	Yes (linear relationship between 22-30)

Family history of diabetes among first-degree relatives	Yes (among first- and second-degree relatives)	
Personal history of GDM	Yes	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	Yes	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	995	995 (same cohort)
Selection		
Number of high risk women		428
Number of low risk women		567
Screening		
Number of women undertook screening	995	995 (same cohort)
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	Random plasma glucose and 50g GCT: screened positive if random plasma glucose ≥6.8mmol/l or 50g GCT ≥ 7.8mmol/l at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	114	N/A
Diagnosis		
Number of women undertook diagnosis	122 (94 women screened positive and 28 women who are a subgroup that OGTT was performed	N/A

	irrespective of the screening tests)	
Time of diagnosis	Within one week of the screening tests	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting 7.0 mmol/l, 2-hour 7.8 mmol/l	Same as universal group
Diagnosis criteria reference	World Health Organisation criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	46/995 (4.6%)	35/995 (3.5%)
Sensitivity	100% (reference)	76.1%=35/46
Specificity	0% (reference)	58.6%=(567-11)/(995-46)
5.3 Other information		
To estimate the proportion of diseased women who were not identified by the screening tests (false negative fraction), in order to correct for verification bias, the study performed OGTT in a subset of 28 women with two negative screening test results. Subsequently, the study used multiple imputation to estimate the results of OGTT in all women in whom no OGTT was performed (24 women out of 873 women who screened negative would be diagnosed as GDM), based on the results of the two screening tests as well as on patient characteristics.		
6. Authors conclusion		
The use of a clinical prediction model is an accurate method to identify women at increased risk for GDM, and could be used to select women for additional testing for GDM.		
7. Reviewer's conclusion		
The author used a clinical prediction model for selecting high risk women for GDM, which could be more accurate than the use of absence/presence of risk factors. The author concluded selective screening by the prediction model could be used, however, the sensitivity of 76% means it could still miss 24% of GDM women, the conclusion needs to be more carefully considered.		

(14) Data extraction result for Williams *et al.* (1999)

1. Study details	
Study ID: 21 **First author surname:** Williams **Year of publication:** 1999 **Country and city:** USA, Ann Arbor (Michigan) **Study design:** Retrospective cohort study **Study setting:** University of Michigan **Number of centres:** 1 **Inclusion and exclusion criteria**: All deliveries at the University of Michigan during the study period **Time of study:** Between 1987 and 1997 **Funding:** Information not available	
2. Aim of the study	
This study aims to estimate the percentage of pregnant women who would not be screened and the percentage of women with GDM who would possibly remain undiagnosed if the American Diabetes Association's (ADA's) new selective screening recommendations are used rather than universal screening for GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	25118
Number of participants consented/included	25118
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	25118
Characteristics of participants	
Mean age (years)	N/A

Mean BMI (kg/m^2)	N/A	
Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	N/A	
4. Selection criteria for high-risk women		
Reference of criteria	The ADA recommendation (1997)	
Type of criteria	Risk factor ≥1	
Number of risk factors used	4	
Maternal age	Yes (≥25)	
Obesity	Yes (BMI≥27, defined by the study)	
Family history of diabetes among first-degree relatives	Yes (among first- and second-degree relatives)	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	Yes (Hispanic, Native American, Asian, African-American)	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	25118	25118 (same cohort)
Selection		
Number of high risk women		N/A
Number of low risk women		N/A
Screening		
Number of women undertook screening	25118	25118 (same cohort)
Time of screening	24-28 weeks of gestation	Same as universal group

Screening test	50g GCT: ≥140mg/dl at 1 hour	Same as universal group
Screening criteria reference	The Second International Workshop-Conference on Gestational Diabetes Mellitus	Same as universal group
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	N/A	Same as universal group
Diagnosis test	100g OGTT (two or more values above thresholds): fasting 105 mg/dl, 1-hour 190 mg/dl, 2-hour 165 mg/dl, 3-hour 145 mg/dl	Same as universal group
Diagnosis criteria reference	National Diabetes Data Group criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	200/25118 (0.8%)	192/25118 (0.8%)
Sensitivity	100% (reference)	96.0%=192/200
Specificity	0% (reference)	11.1% (estimated from a random sample of 171 women)
5.3 Other information		
The author explained why the incidence of GDM in the study appeared low—that's because the population of women with GDM that they studied and the population delivering at the University of Michigan were not identical. Many women who delivered at their institution received care, including testing for GDM, at off-site clinics, and records from those clinics were not available for review.		
6. Authors conclusion		
If the new ADA selective screening recommendations are used, few women with GDM will be missed (4%) but approximately 90% of pregnant women will still need to be screened for GDM.		

7. Reviewer's conclusion
The author did not conclude directly whether they recommend the ADA selective screening approach or not. The selective screening could only exempt 11.1% of the women from screening, thus from the reviewer's view it did not offer much over universal screening. It is also worthwhile to notice that the specificity of 11.1% was estimated from a random sample of 171 women, not from the whole cohort of 25118 women

(15) Data extraction result for Zoller *et al.* (1988)

1. Study details
Study ID: 22 **First author surname:** Zoller **Year of publication:** 1988 **Country and city:** USA, Rockford **Study design:** Prospective cohort study **Study setting:** Teaching hospitals **Number of centres:** Information not available **Inclusion and exclusion criteria**: All patients entering the obstetrical clinics at their teaching hospitals during the study period **Time of study:** Between August 1, 1984, and July 30, 1985 **Funding:** Information not available
2. Aim of the study
This study aims to determine the validity of screening all patients and to determine the cost effectiveness of such a screening program.
3. Participants
Number of participants (One cohort of universal screening)

Number of participants invited/considered	363
Number of participants consented/included	363
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	363
Characteristics of participants	
Mean age (years)	21.3 (13-43)
Mean BMI (kg/m^2)	N/A
Ethnicity	59% white and 41% non-white
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	Reference: Gabbe SG (1980). Effects of identifying a high risk population. *Diabetes Care*. 3: 486-8.
Type of criteria	Risk factor ≥1
Number of risk factors used	8 (10 if 'history of adverse obstetric outcome' is deemed as three RFs)
Maternal age	No
Obesity	Yes (weight ≥90.72kg)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes
History of adverse obstetric outcome	Yes (stillbirth, precious infant with congenital malformation, history of three or more spontaneous abortions)
Member of an ethnic or racial group with a high prevalence of GDM	No
Others	Glucosuria
	Polyhydramnios
	Intrauterine growth consistent with large gestational aged infant
5. Comparison of universal versus selective screening	

	Universal screening	Selective screening
Number of participants	363	363 (same cohort)
Selection		
Number of high risk women		140
Number of low risk women		223
Screening		
Number of women undertook screening	363	363 (same cohort)
Time of screening	24-28 weeks of gestation	Same as universal group
Screening test	50g GCT: ≥140mg/dl at 1 hour	Same as universal group
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	52	N/A
Diagnosis		
Number of women undertook diagnosis	52	N/A
Time of diagnosis	At a later time after the screening test	Same as universal group
Diagnosis test	100g OGTT	Same as universal group
Diagnosis criteria reference	O'Sullivan's diagnostic criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	10/363 (2.8%)	4/363 (1.1%)
Sensitivity	100% (reference)	40%=4/10
Specificity	0% (reference)	61.5%=(223-6)/(363-10)
5.3 Other information		
For universal screening, the total cost for screening and follow-up was $5,190.00. The average cost per patient screened was $14.30, and the cost per case of GDM diagnosed was $519.00.		
6. Authors conclusion		

In order to identify GDM, all pregnant patients must be screened. Universal screening was found to be simple and cost effective.
7. Reviewer's conclusion
The sensitivity and specificity of selective screening was calculated by the reviewer from the data, the author did not calculate them directly. It is rationale for the author to recommend universal screening from the fact that there was no significant difference between the risk group and non-risk group for the number of abnormal OGTTs. The author concluded universal screening was cost-effective, however, it is hard to conclude this only with the average cost per patient screened and cost per case of GDM diagnosed for the universal screening approach.

Appendix 5.2 Study characteristics of the 13 effectiveness studies with a one-step test

(16) Data extraction result for Capula *et al.* (2013)

1. Study details
Study ID: 23 **First author surname:** Capula **Year of publication:** 2013 **Country and city:** Italy, Catanzaro **Study design:** Retrospective cohort study **Study setting:** the Complex Operative Structure of Endocrinology-Diabetology, Pugliese-Ciaccio Hospital, Catanzaro, Calabria, Italy **Number of centres:** 1 **Inclusion and exclusion criteria**: All consecutive pregnant women attending the clinic during the study period were included. Women with pre-existing type 1 or type 2 DM, as defined by ADA criteria, with active chronic systemic disease, and with multifetal gestation were excluded. **Time of study:** Between May 2010 and December 2012 **Funding:** The study did not receive any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

2. Aim of the study	
This study aims to determine the effectiveness of the recent Italian guidelines, which excludes women <35 years old, without risk factors from GDM screening, with respect to the IADPSG criteria.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	2448
Number of participants consented/included	2448
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	2448
Characteristics of participants	
Mean age (years)	30.8 ±3.2 for GDM women; 29.3 ±3.5 for non-GDM women
Mean BMI (kg/m^2)	22.8 ±1.9 for GDM women; 21.4 ±2.0 for non-GDM women
Ethnicity	All Caucasian
Participants with a family history of diabetes among first-degree relatives	N/A
4. Selection criteria for high-risk women	
Reference of criteria	New Italy guidelines for GDM screening (July, 2011)
Type of criteria	Risk factor ≥1
Number of risk factors used	6
Maternal age	Yes (≥35)
Obesity	Yes (BMI ≥25)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes
History of adverse obstetric outcome	No
Member of an ethnic or racial group with a high prevalence of GDM	Yes

Others	No	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	2448	2448 (same cohort)
Selection		
Number of high risk women		1910
Number of low risk women		538
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1910	1910
Time of diagnosis	24-28 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT	Same as universal group
Diagnosis criteria reference	The IADPSG criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	674/2448 (27.5%)	503/2448 (20.5%)
Sensitivity	100% (reference)	74.6%=503/674
Specificity	0% (reference)	20.1%=(538-171)/(2448-674)
5.3 Other information		
The author assessed the maternal and neonatal outcomes in women <35 years without risk factors, and found among these non-risk women, GDM women had significantly more adverse outcomes than non-GDM women in many maternal and neonatal measures. The author also compared the maternal and neonatal outcomes between GDM women with and without risk factors. The lack of differences further indicated the importance of identifying and treating all pregnants with GDM.		
6. Authors conclusion		
Italian recommendations failed to identify 7.0% of women with GDM, under the IADPSG criteria. The risk for adverse hyperglycaemic-related outcomes is		

similar in low-risk and high-risk GDM women. It is suggested to limit OGTT to 1 hour, while extending the test to all pregnant women.
7. Reviewer's conclusion
The prevalence of GDM decreased from 27.5% to 20.5% when screening only high-risk women using the IADPSG diagnosis criteria. The sensitivity is 74.6%, that 25.4% of the GDM women were missed by the selective diagnosis approach. It is rationale for the author to recommend GDM diagnosis among all pregnants. The advantage of the study is that the author additionally compared the adverse hyperglycaemic-related outcomes between GDM and non-GDM in the non-risk group, and between GDM women with and without risk factors. Both results further strengthened the necessity to undertake GDM test for all pregnant women.

(17) Data extraction result for Chong *et al*. (2014)

1. Study details
Study ID: 24 **First author surname:** Chong **Year of publication:** 2014 **Country and city:** Singapore, Singapore **Study design:** Prospective cohort study **Study setting:** the KK Women's and Children's Hospital (KKH) and National University Hospital (NUH) in Singapore **Number of centres:** 2 **Inclusion and exclusion criteria**: Pregnant women 18 years and above, who were in their first trimester were recruited from the two hospitals during the study period. Women who were on chemotherapy, psychotropic drugs or those with type 1 diabetes were excluded. **Time of study:** Between June 2009 and September 2010 **Funding:** The study is supported by the Translational Clinical Research (TCR) Flagship Program on Developmental Pathways to Metabolic Disease funded by the National Research Foundation (NRF). Additional funding is provided by the Singapore Institute for Clinical Sciences – A*STAR.

2. Aim of the study	
This study aims to assess GDM screening approaches in Asian ethnic groups in a single multi-ethnic population, especially to compare the performance of a high-risk screening approach to universal screening for detecting GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1247
Number of participants consented/included	1136 (91.1%; the remaining 111 subjects either declined OGTT or missed their 26–28 weeks clinic visit)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1136
Characteristics of participants	
Mean age (years)	30.7 ±5.1
Mean BMI (kg/m^2)	22.7 ±4.4
Ethnicity	56.7% Chinese, 25.5% Malay and 17.8% Indian
Participants with a family history of diabetes among first-degree relatives	29.4%
4. Selection criteria for high-risk women	
Reference of criteria	UK NICE guideline
Type of criteria	Risk factor ≥1
Number of risk factors used	5
Maternal age	No
Obesity	Yes (BMI >30)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes (weight ≥4.5kg)
History of adverse obstetric outcome	No

Member of an ethnic or racial group with a high prevalence of GDM	Yes (South Asians, black Caribbean, Middle Eastern. South Asian includes India, Pakistan and Bangladesh)	
Others	No	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1136	1136 (same cohort)
Selection		
Number of high risk women		496
Number of low risk women		640
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1136	496
Time of diagnosis	26-28 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting ≥7.0 mmol/l, 2-hour ≥7.8 mmol/l (blood glucose levels were only collected twice to minimize subject burden)	Same as universal group
Diagnosis criteria reference	1999 WHO criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	215/1136 (18.9%)	111/1136 (9.8%)
Sensitivity	100% (reference)	51.6%=111/215
Specificity	0% (reference)	58.2%=(640-104)/(1136-215)
5.3 Other information		
For women of three ethnic groups, 48.4% (104/215) GDM cases would be missed by selective screening. In detail, 66.7% (90/135) Chinese GDM women, 40.0% (14/35) Malay GDM women, and none Indian GDM women would be missed by the selective approach. None Indian GDM cases were missed		

because Indian belongs to the high-risk ethnic group and all would be screened by selective approach anyway.
6. Authors conclusion
Risk factors failed to detect half the GDM cases in Asian women. The study suggests that universal screening for GDM should be instituted in the Singapore population, particularly for Chinese and Indian women.
7. Reviewer's conclusion
The selective screening approach based on GDM risk factors would miss nearly half (48.4%) of the GDM cases. It is rationale for the author to recommend universal screening in the Singapore population. The advantage of the study is that it compared the effect of the selective screening approach among the three Asian ethnic groups, and assessed the risk factors for GDM in the three groups.

(18) Data extraction result for Corrado *et al.* (2014)

1. Study details
Study ID: 25 **First author surname:** Corrado **Year of publication:** 2014 **Country and city:** Italy, Messina **Study design:** Retrospective cohort study **Study setting:** the Clinic of Diabetes and Pregnancy of the University of Messina **Number of centres:** 1 **Inclusion and exclusion criteria**: Consecutive pregnant women referred to the clinic during the study period were included. Women with pre-pregnancy diabetes were excluded. **Time of study:** Between 1 May 2010 and 31 October 2011 **Funding:** Information not available

2. Aim of the study	
This study aims to compare in the population the universal screening test recommended by the IADPSG panel and the ADA versus the selective screening proposed by the UK NICE but modified by the Italian National Institute of Health.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1028
Number of participants consented/included	1015 (13 were excluded: 12 had no complete information on the RFs or the first-trimester FPG values, one had pre-pregnancy diabetes)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1015
Characteristics of participants	
Mean age (years)	32.0 ±4.8 for GDM women; 30.5 ±5.4 for non-GDM women
Mean BMI (kg/m^2)	25.7 ±4.7 for GDM women; 23.7 ±6.8 for non-GDM women
Ethnicity	All Caucasian
Participants with a family history of diabetes among first-degree relatives	40.7% for GDM women; 22.1% for non-GDM women
4. Selection criteria for high-risk women	
Reference of criteria	Selective screening proposed by the UK NICE but modified by the Italian National Institute of Health
Type of criteria	Risk factor ≥1
Number of risk factors used	5
Maternal age	Yes (≥35)
Obesity	Yes (BMI ≥25)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes (birth weight >4500g)

History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	No	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1015	1015 (same cohort)
Selection		
Number of high risk women		591
Number of low risk women		424
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1015	591
Time of diagnosis	24-28 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting ≥92 mg/dl, 1-hour≥180 mg/dl, 2-hour ≥153 mg/dl	Same as universal group
Diagnosis criteria reference	IADPSG and ADA recommendation	Same as universal group
Number of women diagnosed positive-GDM women (%)	113/1015 (11.3%)	87/1015 (8.6%)
Sensitivity	100% (reference)	77.0%=87/113
Specificity	0% (reference)	44.1%=(424-26)/(1015-113)
5.3 Other information		
The author assessed the associations between GDM and the risk factors used in the Italian Institute of Health guidelines. Except maternal age (≥35), all other risk factors were associated with developing GDM. The highest predictive value was confirmed for previous GDM.		
6. Authors conclusion		
More information on the clinical impact of selective screening could be obtained by a strict analysis of treatment, perinatal outcome and follow-up of an adequate sample size of ‘‘missed’’ GDM.		

7. Reviewer's conclusion
The study found that selective screening using the risk factors stated in the Italian Institute of Health guidelines would exempt 44.1% women from screening, while detect 77% of GDM cases (23% would be missed). Normally it would be concluded that selective screening was not recommended. However, the author was very cautious and concluded that only future study of an adequate sample size examining treatment, perinatal outcome and follow-up of the ‘‘missed’’ GDM may answer the question of selective or universal screening. This ‘pending’ conclusion is rationale.

(19) Data extraction result for Cosson *et al.* (2013)

1. Study details
Study ID: 04 **First author surname:** Cosson **Year of publication:** 2013 **Country and city:** France, Paris **Study design:** Retrospective cohort study **Study setting:** Obstetrics department of university hospital (Jean Verdier Hospital) **Number of centres:** 1 **Inclusion and exclusion criteria**: Pregnant women with no known diabetes and with all risk factors known **Time of study:** Between 2002 and 2010 **Funding:** Information not available
2. Aim of the study
This study aims to evaluate a selective screening strategy for GDM based on the presence of risk factors.
3. Participants
Number of participants (One cohort of universal screening)

Number of participants invited/considered	18775	
Number of participants consented/included	18775	
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	87.5% of 18775 women (it is estimated from 2011's data that nearly 12.5% women didn't undertaking the screening)	
Characteristics of participants		
Mean age (years)	29.7 ±5.8	
Mean BMI (kg/m^2)	24.1 ±4.9	
Ethnicity	Multi-ethnicity: Europe (29.1%), North Africa (27.8%) and sub-Saharian Africa (20.8%), and Pakistan, India, and Sri Lanka (5.1%).	
Participants with a family history of diabetes among first-degree relatives	20.4%	
4. Selection criteria for high-risk women		
Reference of criteria	Expert panel for French guidelines	
Type of criteria	Risk factor ≥1	
Number of risk factors used	5	
Maternal age	Yes (≥35)	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes (details N/A)	
History of adverse obstetric outcome	None	
Member of an ethnic or racial group with a high prevalence of GDM	None	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	18775	18775 (same cohort)
Selection		
Number of high risk women		10975
Number of low risk women		7800

Screening		
Number of women undertook screening	None (undertook OGTT directly)	Same as universal group
Time of screening	None	Same as universal group
Screening test	None	Same as universal group
Screening criteria reference	None	Same as universal group
Number of women screened positive (%)	None	Same as universal group
Diagnosis		
Number of women undertook diagnosis	18775	10975
Time of diagnosis	15 weeks of gestation for high-risk women; 24-28 weeks of gestation for all women not diagnosed as GDM at 15 weeks	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting 5.3 mmol/l, 2-hour 7.8 mmol/l	Same as universal group
Diagnosis criteria reference	The fasting value was based on previous French recommendations; the 2-hour value was based on WHO criteria.	Same as universal group
Number of women diagnosed positive-GDM women (%) (GDM prevalence)	2710/18775(14.4%)	1770/18775 (9.4%)
Sensitivity	100% (reference)	65.3%=1770/2710
Specificity	0% (reference)	42.7%=(7800-940)/(18775-2710)
5.3 Other information		
None.		
6. Authors conclusion		
The selective screening would miss one-third of the women with GDM who, even without risk factors, had more events than women without GDM. This study stands against the present selective screening currently proposed in the French guidelines.		
7. Reviewer's conclusion		

The estimated 12.5% women who did not undertake OGTT test might affect the accuracy of the study results of GDM prevalence, sensitivity, and specificity. Based on a sensitivity of 65.3%, it is rationale for the authors to stand against the selective screening approach.

(20) Data extraction result for Jensen *et al.* (2003)

1. Study details
Study ID: 11 **First author surname:** Jensen **Year of publication:** 2003 **Country and city:** Denmark, Copenhagen/ Aarhus/ Odense **Study design:** Prospective cohort study **Study setting:** Four Danish centres: the University Hospitals of Copenhagen (Copenhagen County Hospital and Rigshospitalet), Aarhus, and Odense. **Number of centres:** 4 **Inclusion and exclusion criteria**: Pregnant women who consecutively registered during the study period were included. Exclusion criteria were as follows: pre-existing diabetes mellitus, age younger than 18 years, delivery or migration before 30 weeks, and first booking later than 30 weeks of gestation, as well as women with incomplete data. **Time of study:** Between 1999 to 2000 **Funding:** The study was supported by the Faculty of Health Sciences, University of Southern Denmark, the NOVO Foundation, the Danish Diabetes Association, Handelsgartner Ove Villiam Buhl Olesen og ægtefælle fru E. Buhl Olesens Mindelegat, Direktør Ib Henriksens fond, Poul og Erna Sehested Hansens fond, Fonden til Lægevidenskabens Fremme, the Danish Medical Research Council, and the Grant Committee of the Consultancy, Odense University Hospital.
2. Aim of the study

This study aims to evaluate a simple screening model for GDM on the basis of five pre-defined clinical risk indicators.		
3. Participants		
Number of participants (One cohort of universal screening)		
Number of participants invited/considered	5235	
Number of participants consented/included	2292 (For the other 2943 women, 326 did not speak Danish, 377 could not be contacted despite repeated efforts, and 2240 declined to undergo testing)	
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	2292	
Characteristics of participants		
Mean age (years)	30.0 (27.0-33.4)	
Mean BMI (kg/m^2)	22.7 (20.7-25.4)	
Ethnicity	81% white, 19% others	
Participants with a family history of diabetes among first-degree relatives	14.3%	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	5	
Maternal age	No	
Obesity	Yes (BMI ≥27)	
Family history of diabetes among first-degree relatives	Yes (among parents, children, grandparents, and siblings)	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Glucosuria: +2 or greater on a BM-Test Strip (Boehringer Mannheim, Germany) (equivalent to 5.6 mmol/L)	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**

Number of participants	2292	2292 (same cohort)
Selection		
Number of high risk women		1898
Number of low risk women		3337
Screening		
Number of women undertook screening	No screening test was performed	No screening test was performed
Time of screening	N/A	N/A
Screening test	N/A	N/A
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	2292	1468
Time of diagnosis	Performed in women with either previous GDM or more than one risk factors, and during 28-32 weeks of gestation for all women with risk factors unless already diagnosed with GDM and also in non-risk group	Same as universal group
Diagnosis test	75g OGTT(one or more values above thresholds): fasting 6.1 mmol/l (111 mg/dl), 2-hour 9.0 mmol/l (164 mg/dl)	Same as universal group
Diagnosis criteria reference	Fasting thresholds referred to the World Health Organization criteria for diabetes outside pregnancy. 2-hour thresholds referred to the Diabetic Pregnancy Study Group of the European Association for the Study of Diabetes (DPSG)	Same as universal group
Number of women diagnosed positive-GDM	124/5235 (2.4%)	100/5235 (1.9%)

women (%)		
Sensitivity	100% (reference)	80.6%=100/124
Specificity	0% (reference)	64.8%=(3337-24)/(5235-124)
5.3 Other information		
The prevalence of GDM and the sensitivity were calculated based on the diagnosed GDM cases plus the expected GDM cases from the 2943 (56%) women who did not undergo OGTT.		
6. Authors conclusion		
Under ideal conditions, sensitivity of the selective screening model was comparable with universal screening. The model could avoid two thirds of all pregnant women from the screening and diagnostic testing.		
7. Reviewer's conclusion		
In the study, 56% of participants didn't undergo OGTT, which could cause potential bias to the estimated GDM prevalence and sensitivity. Different from other studies, the risk factors used for selective screening included glucosuria but did not include maternal age. The selection of high risk women was deemed as screening, followed by OGTT diagnosis test. No screening test (such as 50g GCT) was performed.		

(21) Data extraction result for Moses *et al.* (1995)

1. Study details
Study ID: 15 **First author surname:** Moses **Year of publication:** 1995 **Country and city:** Australia, Illawarra area of New South Wales (near the city of Wollongong) **Study design:** Prospective cohort study **Study setting:** The prenatal clinics at Wollongong Hospital **Number of centres:** 1

Inclusion and exclusion criteria: All women attending the prenatal clinic at Wollongong hospital. **Time of study:** Between January 1993 and June 1994 **Funding:** The study was funded by the Illawarra Area Health Service	
2. Aim of the study	
This study aims to determine what proportion of women with GDM by the Australasian Diabetes in Pregnancy Society (ADIPS) criteria would be missed if selective screening as used.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1209
Number of participants consented/included	1185 (98.0%)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1161 (For the other 24 women, data about the BMI were absent in 19 women and the family history was not recorded in 5 women)
Characteristics of participants	
Mean age (years)	26.7%≥ 30 years old
Mean BMI (kg/m^2)	15.1%≥ 30
Ethnicity	N/A
Participants with a family history of diabetes among first-degree relatives	15.3%
4. Selection criteria for high-risk women	
Reference of criteria	N/A
Type of criteria	Risk factor ≥1
Number of risk factors used	3
Maternal age	Yes (≥30)
Obesity	Yes (BMI ≥30)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	No
A prior macrosomic infant	No
History of adverse obstetric outcome	No

Member of an ethnic or racial group with a high prevalence of GDM	No
Others	None

5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1185	1185 (same cohort)
Selection		
Number of high risk women		543
Number of low risk women		642
Screening		
Number of women undertook screening	No screening test was performed	No screening test was performed
Time of screening	N/A	N/A
Screening test	N/A	N/A
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	706 (59.6% of the 1185 women undertook the OGTT test)	N/A
Time of diagnosis	At the beginning of the third trimester	Same as universal group
Diagnosis test	75g OGTT: 2-hour ≥ 8.0 mmol/l (in the first 6 months of the study period); fasting≥ 5.5 mmol/l and/or 2-hour ≥ 8.0 mmol/l (in the following 12 months of the study period)	Same as universal group
Diagnosis criteria reference	The Australasian Diabetes in Pregnancy Society (ADIPS) criteria	Same as universal group

Number of women diagnosed positive-GDM women (%)	79/1185 (6.7%)	48/1185 (4.1%)
Sensitivity	100% (reference)	60.8%=48/79
Specificity	0% (reference)	55.2%=(642-31)/(1185-79)
5.3 Other information		
None.		
6. Authors conclusion		
This study supports the ADIF'S recommendation that there should be universal testing.		
7. Reviewer's conclusion		
No screening test was performed, GDM was diagnosed directly by the 75g OGTT test. In the study, only 706 (59.6%) of the 1185 participants underwent the diagnosis test, which could lead to potential bias in the estimate of GDM prevalence and the sensitivity of selective screening.		

(22) Data extraction result for Moses *et al.* (1998)

1. Study details
Study ID: 16 **First author surname:** Moses **Year of publication:** 1998 **Country and city:** Australia, Illawarra area of New South Wales (near the city of Wollongong) **Study design:** Retrospective cohort study **Study setting:** The prenatal clinics at Wollongong and Shellharbour hospitals **Number of centres:** 2 **Inclusion and exclusion criteria**: All pregnant women attending the prenatal clinics at Wollongong and Shellharbour hospitals. Only the data from

singleton pregnancies have been included **Time of study:** Between January 1993 and June 1994 **Funding:** Information not available	
2. Aim of the study	
This study aims to determine the GDM prevalence in women with low-risk factors and to see if the pregnancy outcomes of women with GDM from a low-risk group were diff e rent from the outcomes of other women with GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	2907
Number of participants consented/included	2907
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	2907
Characteristics of participants	
Mean age (years)	27.0 ±5.0 (for Caucasian women)
Mean BMI (kg/m^2)	24.1 ±5.9 (for Caucasian women)
Ethnicity	91.1% Caucasian (72.7% Australian, 18.4% European), 0.7% Aboriginal, 0.7% Pacific Islander, 3.1% Asian, 4.4% others
Participants with a family history of diabetes among first-degree relatives	14.6%
4. Selection criteria for high-risk women	
Reference of criteria	N/A
Type of criteria	Risk factor ≥1
Number of risk factors used	3
Maternal age	Yes (≥25)
Obesity	Yes (BMI ≥25)
Family history of diabetes among first-degree relatives	No

Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	Yes (non-Caucasian ethnic origin)	
Others	None	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	2907	2907 (same cohort)
Selection		
Number of high risk women		2334
Number of low risk women		573
Screening		
Number of women undertook screening	No screening test was performed	No screening test was performed
Time of screening	N/A	N/A
Screening test	N/A	N/A
Screening criteria reference	N/A	N/A
Number of women screened positive (%)	N/A	N/A
Diagnosis		
Number of women undertook diagnosis	N/A	N/A
Time of diagnosis	At the beginning of the third trimester	Same as universal group
Diagnosis test	75g OGTT: 2-hour ≥ 8.0 mmol/l	Same as universal group
Diagnosis criteria reference	The Australasian Diabetes in Pregnancy Society	Same as universal group

	(ADIPS) criteria	
Number of women diagnosed positive-GDM women (%)	183/2907 (6.3%)	167/2907 (5.7%)
Sensitivity	100% (reference)	91.3%=167/183
Specificity	0% (reference)	20.4%=(573-16)/(2907-183)
5.3 Other information		
The study additionally assessed the figures of specificity and sensitivity using different selection and diagnosis criteria. If the BMI thresholds changed from ＜25 to ＜ 27, then the low risk women would increase from 573 to 652, and the GDM diagnosed would decrease from 167 to 149, the new specificity and sensitivity would be 22.4% (↑) and 81.4% (↓), respectively. If the 2-hour glucose level of GDM diagnosis changed from ≥ 8.0 mmol/l to ≥ 9.0 mmol/l, then the GDM cases would decrease from 167 to 75, and 6 cases would be from the 573 low risk women, the new sensitivity would be 92% (similar to the old sensitivity of 91.3%).		
6. Authors conclusion		
The pregnancy outcomes of women with GDM from a low-risk group are similar to the outcomes of other women with GDM. Selective screening needs further evaluation in different populations before it can be endorsed.		
7. Reviewer's conclusion		
The study included 2907 women from Wollongong and Shellharbour hospitals, but didn't report the number of women who actually underwent the OGTT diagnosis test. From a precious study of the author (Moses *et al.,* 1995) including 1185 women from Wollongong hospital of the same study period, only 706 (59.6%) of the 1185 women undertook the OGTT test. The non-compliance with the test could cause potential bias in the estimate of GDM prevalence and sensitivity. The advantage of the study is that the author additionally assessed the figures of specificity and sensitivity using different selection and diagnosis criteria (BMI thresholds changed from ＜25 to ＜ 27; the 2-hour glucose level of GDM diagnosis changed from ≥ 8.0 mmol/l to ≥ 9.0 mmol/l), and found the sensitivity and specificity would change according to the different criteria used.		

(23) Data extraction result for Ostlund & Hanson (2003)

1. Study details	
Study ID: 27 **First author surname:** Ostlund & Hanson **Year of publication:** 2003 **Country and city:** Sweden, Orebro **Study design:** Prospective cohort study **Study setting:** Maternal health centres and delivery departments in Orebro County **Number of centres:** Information not available **Inclusion and exclusion criteria**: All pregnant non-diabetic women attending maternal health care in Orebro County in Sweden during the study period **Time of study:** Between 1 July 1994 and 30 June 1996 **Funding:** The study was supported by grants from the General Maternity Hospital Foundation and Research Committee of Orebro County Council	
2. Aim of the study	
This study aims to determine prevalence of GDM, and the value of traditional anamnestic risk factors for predicting GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	4918
Number of participants consented/included	3616 (73.5%) agreed to perform the OGTT
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	3616
Characteristics of participants	
Mean age (years)	27.9 ±4.8
Mean BMI (kg/m^2)	23.8 ±4.1
Ethnicity	11.2% Nonnordic origin

Participants with a family history of diabetes among first-degree relatives	N/A	
4. Selection criteria for high-risk women		
Reference of criteria	Recommendation from the Swedish National Board of Health and Welfare in 1997	
Type of criteria	Risk factor ≥1	
Number of risk factors used	4	
Maternal age	No	
Obesity	Yes (weight ≥90 kg)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	No	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	3616	3616 (same cohort)
Selection		
Number of high risk women		571
Number of low risk women		3045
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	3616	571
Time of diagnosis	28-32 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting ≥6.7 mmol/l, 2-hour ≥9.0 mmol/l. GDM include IGT and DM. IGT was diagnosed if	Same as universal group

	fasting <6.7 mmol/l and 2-hour between 9.0–11.0 mmol/l; DM was diagnosed if fasting ≥6.7 mmol/l or 2-hour ≥11.1 mmol/l.	
Diagnosis criteria reference	The World Health Organization (1980)	Same as universal group
Number of women diagnosed positive-GDM women (%)	61/3616 (1.7%)	29/3616 (1.1%)
Sensitivity	100% (reference)	47.5%=29/61
Specificity	0% (reference)	84.8%=(3045-32)/(3616-61)
5.3 Other information		
Random B-glucose was analysed every fourth to six week. If random B-glucose was ≥9.0 mmol/l an OGTT was carried out immediately. If this was in early pregnancy, an OGTT was repeated in 28–32 weeks of gestation. The sensitivity and specificity using the four risk factors are 47.5% and 84.8%, respectively. Adding ethnicity (Nonnordic) to these risk factors increased the sensitivity to 60.7% and decreased the specificity to 74.8%. While further adding maternal age (≥25 years) increased the sensitivity to 93.4%, however, the specificity decreased to 20.3%.		
6. Authors conclusion		
Using traditional risk factors as an indicator to perform an OGTT gives a low sensitivity to detect GDM and even DM especially among primiparas.		
7. Reviewer's conclusion		
The study analysed the sensitivity and specificity under three different selection criteria. Using the four traditional risk factors gave a high specificity (84.8%) but low sensitivity (47.5%). Adding ethnicity as a risk factor didn't change much (74.8% and 60.7%, respectively). While further adding maternal age as a risk factor gave a high sensitivity (93.4%) but unfortunately low specificity (20.3%). It is reasonable for the author to conclude that using traditional risk factors gives a low sensitivity and there is a need for other screening models.		

(24) Data extraction result for Pintaudi *et al.* (2014)

1. Study details	
Study ID: 28 **First author surname:** Pintaudi **Year of publication:** 2014 **Country and city:** Italy, Messina **Study design:** Retrospective cohort study **Study setting:** the Clinic of Diabetes and Pregnancy of the University of Messina **Number of centres:** 1 **Inclusion and exclusion criteria**: Consecutive pregnant women referred to the clinic during the study period were included. Women with pre-pregnancy diabetes were excluded. **Time of study:** Between 1 May 2010 and 31 October 2011 **Funding:** The study did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector	
2. Aim of the study	
This study aims to assess the predictive value of risk factors (RFs) for GDM established by selective screening (SS) and to identify subgroups of women at a higher risk of developing GDM.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1028
Number of participants consented/included	1015 (13 were excluded: 12 had no complete information on the RFs or the first-trimester FPG values, one had pre-pregnancy diabetes)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1015
Characteristics of participants	

Mean age (years)	32.0 ±4.8 for GDM women; 30.5 ±5.4 for non-GDM women	
Mean BMI (kg/m^2)	25.7 ±4.7 for GDM women; 23.7 ±6.8 for non-GDM women	
Ethnicity	All Caucasian	
Participants with a family history of diabetes among first-degree relatives	40.7% for GDM women; 22.1% for non-GDM women	
4. Selection criteria for high-risk women		
Reference of criteria	RECPAM model: risk factors identified by the RECursive Partitioning and AMalgamation (RECPAM) method	
Type of criteria	Risk factor ≥1	
Number of risk factors used	3	
Maternal age	No	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	FPG >4.4 mmol/l	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1015	1015 (same cohort)
Selection		
Number of high risk women		641
Number of low risk women		374
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1015	641

Time of diagnosis	24-28 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (one or more values above thresholds): fasting ≥5.1 mmol/l, 1-hour≥10.0 mmol/l, 2-hour ≥8.5 mmol/l.	Same as universal group
Diagnosis criteria reference	IADPSG recommendation	Same as universal group
Number of women diagnosed positive-GDM women (%)	113/1015 (11.3%)	101/1015 (10.0%)
Sensitivity	100% (reference)	89.0%=101/113
Specificity	0% (reference)	40.1%=(374-12)/(1015-113)
5.3 Other information		
The author conducted the RECPAM analysis to identify women at different risks of developing GDM. The study compared the effectiveness of selective screening using RECPAM model and SS criteria. The application of SS criteria would result in the execution of OGTT in 591 (58.3%) of women, and 26 (23.0%) GDM cases would be missed due to the absence of any RF. The RECPAM model, however, would reduce by over 50% (23.0 vs 10.6%) the number of undiagnosed GDM cases when compared with the current SS approach, at the expense of 50 additional OGTTs required. The RFs used in the SS criteria are: age ≥35 years, BMI ≥25, FPG values between 5.6 and 6.9 mmol/l during pre-pregnancy or in the first trimester of pregnancy, previous GDM, previous macrosomia (≥4500g), family history of diabetes (first-degree relative with diabetes), and origin of family from areas with a high prevalence of diabetes.		
6. Authors conclusion		
A selective screening approach based on our RECPAM model results in a significant reduction in the number of undetected GDM cases compared with the current selective screening approach using SS critieria.		
7. Reviewer's conclusion		
The selective screening approach using the RECAPM model would exempt 40.1% pregnant women from the OCTT and still diagnose 89.0% of GDM cases. Compared with the selective screening approach using traditional RFs, it increased the sensitivity from 77.0% (87/113) to 89.0% at the cost of only 50 more OGTTs. It is rationale for the author to recommend the RECAPM model. The advantage of the study is that the author employed the RECAPM method to identify the RFs significantly associated with GDM in the population, and produced a more accurate selective screening model than the traditional one.		

(25) Data extraction result for Savona-Ventura *et al.* (2013)

1. Study details
Study ID: 29 **First author surname:** Savona-Ventura **Year of publication:** 2013 **Country and city:** 11 Mediterranean countries **Study design:** Prospective cohort study **Study setting:** Centres among 11 Mediterranean countries **Number of centres:** Information not available **Inclusion and exclusion criteria**: Each participating centre from the 11 countries recruited a convenience sample of 75–200 pregnant women attending routine prenatal care during the study period. The participants were not known to have any form of carbohydrate metabolism disorder before their current pregnancy (i.e. T1DM, T2DM, or maturity-onset diabetes of the young). **Time of study:** Between August 1, 2010, and May 31, 2011 **Funding:** The study was funded by a financial grant from the Mediterranean Group for the Study of Diabetes, which is supported by an unrestricted educational grant from Servier.
2. Aim of the study
This study aims to determine whether clinical risk assessment for GDM may preclude the need for universal screening with an OGTT in situations of economic restraint.
3. Participants

Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1368
Number of participants consented/included	1368
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1368

Characteristics of participants		
Mean age (years)	Eastern Mediterranean countries (28.9±6.1), northern Mediterranean countries (30.3±5.3), Maghreb countries (30.2±5.8)	
Mean BMI (kg/m^2)	Maternal BMI during the third trimester: Eastern (28.2±5.3), northern (28.5±4.9), Maghreb countries (27.6±4.7)	
Ethnicity	64 women (4.7%) were born in countries outside the Mediterranean region, 13 of these 64 women had 1 or both parents of Mediterranean origin	
Participants with a family history of diabetes among first-degree relatives	N/A	
4. Selection criteria for high-risk women		
Reference of criteria	From the results of the risk factors analysis in this current study	
Type of criteria	FBG >5.0 mmol/L or maternal obesity with age ≥30	
Number of risk factors used	3	
Maternal age	Yes (≥30)	
Obesity	Yes (pre-pregnancy BMI ≥25 or a 3rd trimester BMI ≥30)	
Family history of diabetes among first-degree relatives	No	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	FBG >5.0 mmol/L	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1368	1368 (same cohort)
Selection		
Number of high risk women		466
Number of low risk women		902

Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1368	466
Time of diagnosis	24-32 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT	Same as universal group
Diagnosis criteria reference	The American Diabetes Association (ADA) criteria (2003)	Same as universal group
Number of women diagnosed positive-GDM women (%)	119/1368 (8.7%)	96/1368 (7.0%)
Sensitivity	100% (reference)	80.7%=96/119
Specificity	0% (reference)	70.4%=(902-23)/(1368-119)
5.3 Other information		
The author calculated the sensitivity and specificity using nine different combinations of risk factors (see Table 2 in the original article). Apart from the reported combination of risk factors, another important combination was FBG >5.0 mmol/L or age ≥30 years or diastolic BP ≥80 mmHg. Under this combination, the sensitivity and specificity was 96.6% and 34.5%, respectively.		
6. Authors conclusion		
Use of a composite model to prescreen women for GDM risk may reduce the need for universal screening with the OGTT among centres facing health-cost pressures.		
7. Reviewer's conclusion		
The figures of specificity in this study were re-calculated by the reviewer using the definition of specificity in the review. Although the author assessed the sensitivity and specificity under nine different combinations of risk factors, the author's conclusion was made based on the first combination which was FBG >5.0 mmol/L or maternal obesity with age ≥30 (sensitivity and specificity was 80.7% and 65.9%, respectively). It is rationale for the author to recommend selective screening under this combination for the areas of economic restraint.		

(26) Data extraction result for Shamsuddin *et al.* (2001)

1. Study details	
Study ID: 30 **First author surname:** Shamsuddin **Year of publication:** 2001 **Country and city:** Malaysia, Cheras **Study design:** Prospective cohort study (reported as cross-sectional study by the author) **Study setting:** Antenatal clinic at the Hospital Universiti Kebangsaan Malaysia (HUKM) **Number of centres:** 1 **Inclusion and exclusion criteria**: Pregnant women at ≥24 weeks' gestation who were attending the antenatal clinic in HUKM. Known cases of diabetes mellitus or GDM already diagnosed elsewhere prior to the first antenatal visit in HUKM were excluded. **Time of study:** Between June 1999 and January 2000 **Funding:** Information not available	
2. Aim of the study	
This study aims to assess the prevalence and association of frequently used screening risk factors for GDM, and to compare the validity and cost of universal screening with risk factor screening.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	896
Number of participants consented/included	835 (93.2% responded)
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	768 (For the other 67 women, laboratory results of 50 mothers could not be traced and another 17 mothers were excluded due to incomplete data)
Characteristics of participants	
Mean age (years)	N/A (0.3% <20; 85% between 20-24; 14.7% ≥35)

Mean BMI (kg/m^2)	N/A	
Ethnicity	58.3% Malays, 35.0% Chinese, 5.0% Indians, 1.7% Others	
Participants with a family history of diabetes among first-degree relatives	N/A	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	9	
Maternal age	Yes (≥35)	
Obesity	Yes (maternal weight ≥80kg)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	Yes	
A prior macrosomic infant	Yes (≥4000g)	
History of adverse obstetric outcome	Yes (birth with congenital anomalies, intrauterine deaths, neonatal death, previous polyhydramnios, spontaneous abortion)	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Urinary tract infection	
	Vaginal discharge and pruritis vulvae	
	Glycosuria	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	768	768 (same cohort)
Selection		
Number of high risk women		513
Number of low risk women		255
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	768	513

Time of diagnosis	≥ 24 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (one value above threshold): 2-hour 7.8mmol/l	Same as universal group
Diagnosis criteria reference	N/A	Same as universal group
Number of women diagnosed positive-GDM women (%)	191/768 (24.9%)	138/1368 (10.1%)
Sensitivity	100% (reference)	72.2%=138/191
Specificity	0% (reference)	35.0%=(255-53)/(768-191)
5.3 Other information		
Among all these traditional risk factors used for the selective screening, only previous history of GDM and maternal age were significantly associated with prevalence of GDM in this population. The study also evaluated the cost of universal and selective screening, which were 12.6 RM and 11.15 RM to identify a case of GDM, respectively.		
6. Authors conclusion		
Risk factor screening scored poorly in predicting GDM. Cost analysis of universal compared with traditional risk factor screening showed a negligible difference. Thus universal screening appears to be the most reliable method of diagnosing GDM.		
7. Reviewer's conclusion		
The specificity in this study was re-calculated by the reviewer using the definition of specificity in the review, which was the proportion of low-risk women who could be exempted from the GDM screening/diagnosis. The selective screening used a series of traditional risk factors but only identified 72.2% of the GDM cases. It is reasonable for the author to recommend universal screening to improve GDM detection in their population.		

(27) Data extraction result for Shirazian *et al.* (2009)

1. Study details	
Study ID: 31 **First author surname:** Shirazian **Year of publication:** 2009 **Country and city:** Iran, Tehran **Study design:** Retrospective cohort study (a secondary analysis on the ongoing prospective study) **Study setting:** Javaheri General Hospital and four private obstetric clinics **Number of centres:** 5 **Inclusion and exclusion criteria**: Pregnant women with the diagnosis of diabetes before pregnancy, presence of comorbid conditions and whose gestational age was more than 28 weeks at first prenatal visit were excluded from the study. **Time of study:** Between May 2005 and May 2008 **Funding:** The study was supported by Islamic Azad University, Tehran Medical Branch grant	
2. Aim of the study	
This study aims to determine the influence of risk factors on incidence of GDM in Iranian population.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	971
Number of participants consented/included	971
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	924 (47 women did not complete OGTT)
Characteristics of participants	
Mean age (years)	N/A (29.2% ≤24; 40.2% between 25-29; 30.6% >30)
Mean BMI (kg/m^2)	N/A (49.4% ≤24.9; 36.6% between 25.0-29.9; 14.0% >30)

Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	22.0%	
4. Selection criteria for high-risk women		
Reference of criteria	Risk factors used was identified from the current study	
Type of criteria	Risk factor ≥1	
Number of risk factors used	3	
Maternal age	Yes (≥25)	
Obesity	Yes (BMI ≥25)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	No	
A prior macrosomic infant	No	
History of adverse obstetric outcome	No	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	No	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	924	924 (same cohort)
Selection		
Number of high risk women		798
Number of low risk women		126
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	924	798
Time of diagnosis	24-28 weeks of gestation	Same as universal group
Diagnosis test	75g OGTT (two or more values above thresholds): fasting 5.3 mmol/l (95 mg/dl), 1-hour	Same as universal group

	10.0 mmol/l (180 mg/dl), 2-hour 8.6 mmol/l (155 mg/dl)	
Diagnosis criteria reference	American Diabetes Association (ADA) criteria	Same as universal group
Number of women diagnosed positive-GDM women (%)	68/924 (7.4%)	67/924 (7.3%)
Sensitivity	100% (reference)	98.5%=67/68
Specificity	0% (reference)	14.6%=(126-1)/(924-68)
5.3 Other information		
The author established a risk score system using the odds ratios of GDM risk factors. Pregnant women were stratified into 6 different score risk groups, scoring from 0 to 6. However, the low-risk group was directly defined as the group who scored 0, without further exploration of the other groups scored 1-6. The sensitivity, specificity, and the conclusion was calculated or made based on this definition of low-risk group.		
6. Authors conclusion		
Age, BMI, and family history of diabetes were independent risk factors in developing GDM. Concerning these factors, we do not miss substantial number of GDM cases with selective screening.		
7. Reviewer's conclusion		
The selective screening approach exempted 14.6% women from the GDM diagnosis and missed 1.5% of GDM women. It is rationale for the author to conclude that selective screening will not miss substantial number of GDM cases in Iranian people using their selection criteria. However, the proportion of GDM women exempted from the GDM diagnosis test was also low. The author didn't make a conclusion about whether or not to recommend selective screening.		

(28) Data extraction result for Wagaarachchi *et al.* (2001)

1. Study details	
Study ID: 32 **First author surname:** Wagaarachchi **Year of publication:** 2001 **Country and city:** Sri Lanka, Colombo **Study design:** Prospective cohort study **Study setting:** Antenatal clinics at the Castle Street Hospital for Women, Colombo **Number of centres:** 1 **Inclusion and exclusion criteria**: Pregnant women attending antenatal clinics at the hospital during the study period were included. Women with pre-existing diabetes were excluded from the study **Time of study:** 22 months (details not available) **Funding:** The study was supported by grants from the Sri Lanka College of Obstetricians and Gynaecologists and Les Laboratories Survier, Novo Nordisk Limited.	
2. Aim of the study	
This study aims to determine what proportion, if any, of women with GDM would be missed if testing was confined to clinical risk factors in an Asian population.	
3. Participants	
Number of participants (One cohort of universal screening)	
Number of participants invited/considered	1096
Number of participants consented/included	1096
Number of participants in the end after those missed/excluded/lost to follow-up at a later stage	1004 (who had both tests performed and completed the study)
Characteristics of participants	

Mean age (years)	N/A (21.6% ≥35)	
Mean BMI (kg/m^2)	N/A (11.1% ≥30)	
Ethnicity	N/A	
Participants with a family history of diabetes among first-degree relatives	13.3%	
4. Selection criteria for high-risk women		
Reference of criteria	N/A	
Type of criteria	Risk factor ≥1	
Number of risk factors used	6	
Maternal age	Yes (≥35)	
Obesity	Yes (BMI ≥30)	
Family history of diabetes among first-degree relatives	Yes	
Personal history of GDM	No	
A prior macrosomic infant	Yes	
History of adverse obstetric outcome	Yes (history of unexplained perinatal loss)	
Member of an ethnic or racial group with a high prevalence of GDM	No	
Others	Presence of grand multiparity	
5. Comparison of universal versus selective screening		
	Universal screening	**Selective screening**
Number of participants	1004	1004 (same cohort)
Selection		
Number of high risk women		461
Number of low risk women		543
Diagnosis (one-step diagnosis only)		
Number of women undertook diagnosis	1001	461
Time of diagnosis	24-28 weeks of gestation	Same as universal group

Diagnosis test	75g OGTT	Same as universal group
Diagnosis criteria reference	The World Health Organisation Criteria (1980)	Same as universal group
Number of women diagnosed positive-GDM women (%)	41/1004 (4.1%)	24/1004 (2.4%)
Sensitivity	100% (reference)	58.5%=24/41
Specificity	0% (reference)	54.6%=(543-17)/(1004-41)
5.3 Other information		
Of the six risk factors used for the selective screening, only maternal age (≥35 years) and obesity (BMI ≥30) were significantly associated with developing GDM.		
6. Authors conclusion		
GDM is common in Asian women, and screening should be performed on every pregnant woman, rather than on selective basis as indicated by risk factors.		
7. Reviewer's conclusion		
In the study, the selective screening approach using six risk factors would exempt about half of the women (54.6%) from the OGTT test, but would also miss around half GDM women (58.5%). It is rationale for the author to recommend universal screening. One point is that among the six risk factors, only two (maternal age and obesity) were found to be significantly associated with GDM. If use only these two risk factors and use different cut-offs (other than 35 for age and 30 for obesity), there is possibility that the conclusion might change due to the altered specificity and sensitivity.		

Appendix 5.3 Study characteristics of the cost-effectiveness study

Data extraction result for Poncet *et al.* (2002)

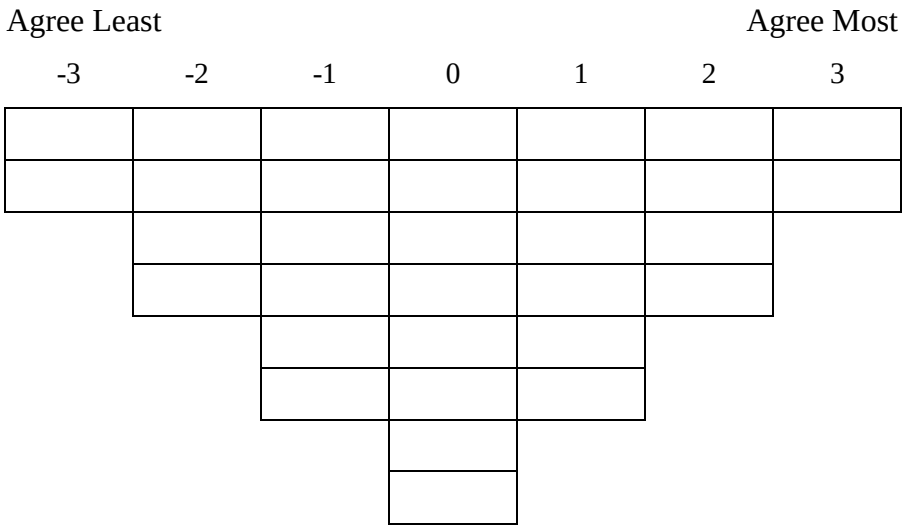
1. Study details	
Study ID: 33 **First author surname:** Poncet **Year of publication:** 2002 **Country:** France **Analytical framework (type of model):** Decision analysis and cost-effectiveness analysis **Perspective:** Health system **Study population**: Pregnant women undertaking GDM screening and diagnosis at hospital **Funding:** The study was supported by grants from the Programme Hospitalier de Recherche Clinique 1997 (PHRC-Health ministry)	
2. Aim of the study	
This study aims to compare three strategies for GDM screening, in terms of cost-effectiveness.	
3. Intervention and comparator	
Intervention	S1: Screening of high-risk pregnant women with 50g GCT followed by 100g OGTT S2: Screening of all pregnant women with 50g GCT followed by 100g OGTT S3: Screening of all pregnant women with a one-step 75g OGTT
Comparator	S0: No screening
Details of screening and diagnosis tests	50g GCT: 1-hour ≥7.2 mmol/l (Carpenter and Coustan criteria) 100g OGTT (two or more values above thresholds): fasting ≥5.3 mmol/l, 1-hour ≥10.0 mmol/l, 2-hour ≥8.6 mmol/l, 3-hour ≥7.8 mmol/l (Carpenter and Coustan criteria) 75g OGTT (one or more values ≥ thresholds): fasting ≥5.5 mmol/l, 2-hour ≥8.0 mmol/l (World Health Organization criteria)

4. Selection criteria for high-risk women	
Reference of criteria	N/A
Type of criteria	Risk factor ≥1
Number of risk factors used	6
Maternal age	Yes (≥35)
Obesity	Yes (BMI ≥27)
Family history of diabetes among first-degree relatives	Yes
Personal history of GDM	Yes
A prior macrosomic infant	Yes (≥ 4000g)
History of adverse obstetric outcome	Yes (pre-eclampsia, foetal death after 3 months of gestation)
Member of an ethnic or racial group with a high prevalence of GDM	No
Others	No
5. Clinical effectiveness	
Effectiveness measures	The outcome measures of effectiveness were macrosomia, prematurity, perinatal mortality, and hypertensive disorders rates.
Source of effectiveness data	The effectiveness estimations were extracted from a literature analysis based on Medline data using relevant keywords. Thirty-eight articles, published between 1973 and 1998 were finally retained and used to calculate the effectiveness.
6. Cost data	
Economic measures	The costs accounted for screening tests, obstetrical cares, management of GDM, if any, delivery cares, and sick leave, starting from the 24th week of gestation till discharge from maternity. The long-term consequences of GDM after delivery were not considered.
Source for resource consumption	To estimate the various costs, they conducted a prospective study of 120 pregnancies between 15

	February and 11 April 1999, in a public hospital of the Rhone-Alpes region (Lyon Sud Hospital-Hospices Civils de Lyon). They relied also on expert opinions and literature data as complements.
Reference of cost (health care system)	The costs were evaluated according to the official French health insurance system (Assurance Maladie) criteria. The costs were estimated using the French key-letters costing system. Hospitalisation costs were evaluated using ISA points (synthetic activity index) per GHM (homogenous group of patients, the French version of diagnosis related group). In 1997, the ISA point value as Euros 1.94. Costing of sick leave was estimated from the daily allowance for sick leave paid by the official French health insurance system.
Currency used	Costs estimated in euros
Years to which costs apply	1997 (Costing up-dating was not necessary because all costs were calculated within 3 months)
7. Cost-effectiveness	
Outcome measures used in economic evaluations	ICER
Modelling summary	A model of decision analysis was used to estimate the cost-effectiveness ratios (CERs) of the three alternative screening strategies versus the strategy of 'no screening'. Based on the CERs, the estimation of the ICERs of each screening strategy with no screening was produced and compared.
Cost-effectiveness outcomes	The ICERs of S1 are 21069.52 (macrosomia), 9953.24 (prematurity), 7871.55 (perinatal mortality), and 28674.90 (hypertension disorders). The ICERs of S2 are 23135.36 (macrosomia), 10965.20 (prematurity), 8663.83 (perinatal mortality), and 31898.74 (hypertension disorders). The ICERs of S3 are 68933.79 (macrosomia), 37320.89 (prematurity), 29444.16 (perinatal mortality), and 94506.04 (hypertension disorders). ICERs of S2 were 1.10-1.11 times of S1. ICERs of S3 were 3.27-3.75 times of S1.
Sensitivity analysis	Sensitivity analysis for the key variables was performed.
6. Other information	
The sensitivity analysis showed that: (1) Changing the values of the events and the values of the main outcome measures within the range found in the	

literature did not change of results of the ICERs; (2) Changing the values of the costs within the ranges [Euros 3811.23-6860.21] changed the ICERs values. Above the thresholds of Euros 5488.16, that is a relative variation of 9.5%, the strategy of screening of all pregnant women with a 75g OGTT (S3) was proved to be the most cost-effective.
7. Authors conclusion
The costs per case prevented reflect a favourable cost-effectiveness ratio for screening of high-risk pregnant women by 50g GCT.
8. Reviewer's conclusion
The ICER, that is cost to obtain one unit of additional effectiveness, was 1.10-1.11 times more expensive of S2, and 3.27-3.75 times more expensive of S3, compared with S1. It is rationale for the author to conclude that S1 (selective screening) was the most cost-effective strategy among the three screening strategies. It is worthwhile to notice that the conclusion was not outstandingly robust, since sensitivity analysis showed changing the values of the costs could change the ICERs and the conclusion (S3 could be the most cost-effective strategy under certain circumstances). It is also worthwhile to notice that the outcome measures used are short-term outcomes before discharge from the hospital. The author also pointed out the future need for cost-effectiveness estimation on a long-term basis.

Appendix 6. Q sorting from −3 (Agree least) to +3 (Agree most) for the 32 Q statements



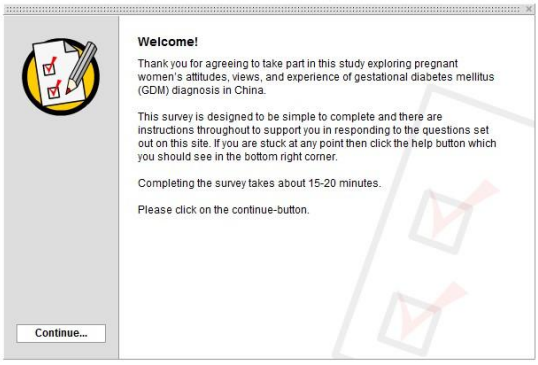
Appendix 7. The 32 Q statements for the Q methodology study

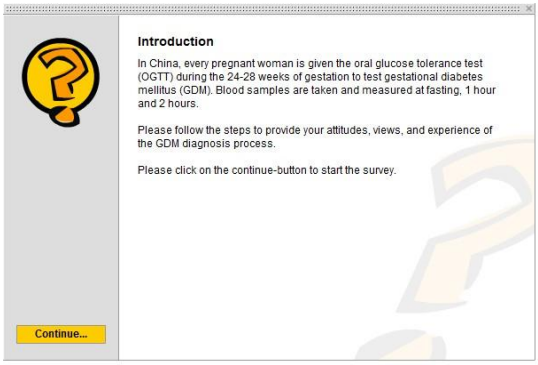
1. The OGTT is convenient
2. Having blood drawn three times within 2 hours OGTT is too much
3. The duration of 2 hours for the OGTT test is too long
4. The time from having the OGTT test to collecting the result was satisfactory
5. I feel it is more convenient to collect my GDM diagnosis result during my next visit to the hospital, rather than waiting for several hours to collect it on that day
6. I prefer face-to-face notification of the GDM diagnosis result when I come to the hospital rather than being telephoned when the result comes out
7. It was easy to find the scan machine for collecting my GDM diagnosis result
8. The GDM diagnosis result sheet was NOT understandable
9. I was satisfied with the doctor's explanation when I gave my GDM diagnosis result sheet to him/her to review
10. I wish I had been given more information about GDM and GDM diagnosis before I had the OGTT
11. I wish I had been given more information about GDM and GDM diagnosis after I had the OGTT

12. I am NOT satisfied with the GDM information I received from the doctors/nurses at the hospital
13. I am satisfied with the GDM information I received from the weekly lectures for pregnant women at the hospital
14. I am NOT satisfied with the GDM information I received from the leaflet/bulletin board at the hospital
15. I would prefer to receive GDM information from weekly lectures for pregnant women at the hospital
16. I would prefer to receive GDM information from an education leaflet provided by the hospital
17. I would NOT prefer to receive GDM information from the hospital bulletin board
18. I would NOT prefer to search for GDM information (from internet, TV, magazines, books, etc.) by myself
19. I would prefer my family (parents, husband/partner) or my friends to search for GDM information for me
20. I would prefer to learn about GDM from talking to women who have or had GDM
21. Testing for GDM is very important and necessary
22. The OGTT should be conducted for all pregnant women
23. I feel it is a burden to undergo the OGTT
24. I was confused about the OGTT
25. I was unhappy with the OGTT
26. I am satisfied with the whole process of being tested for GDM
27. I am satisfied with the support I received throughout the OGTT procedure
28. I was treated with dignity and respect by staff in the hospital throughout the whole OGTT procedure
29. The doctor who gave the OGTT to me was NOT knowledgeable and informative
30. The doctor who did the OGTT listened to my concerns and listened to what I needed
31. Sometimes I felt afraid to ask the hospital staff relevant advice about OGTT and GDM
32. All necessary explanations and advice about OGTT were given by hospital staff

Appendix 8. Screenshots illustration of the FlashQ programme

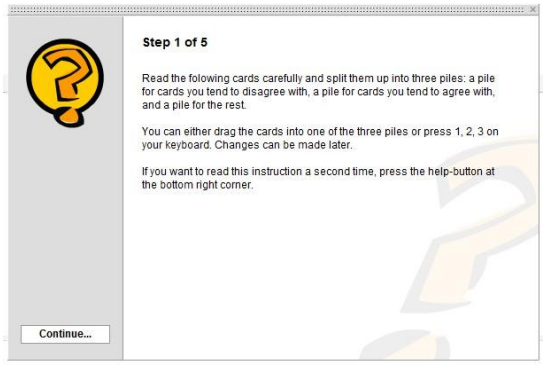
1. Instruction：

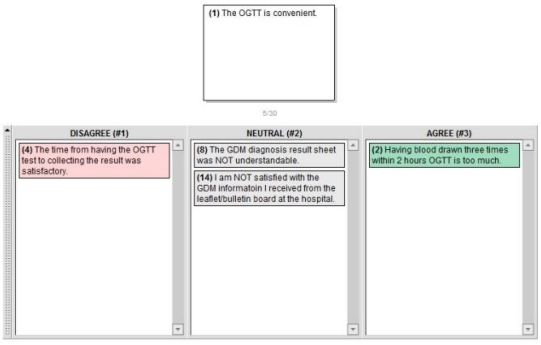




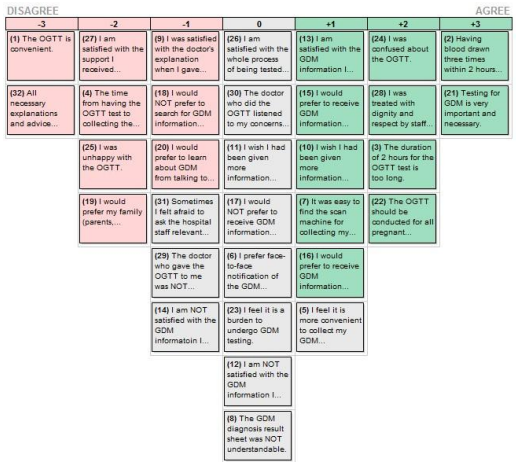
2. Q statements and Q sorting:

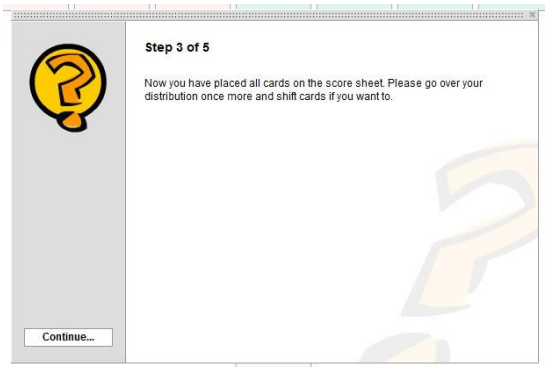
2.1Select 'Disagree', 'Neutral', or 'Agree' for each of the 32 statements



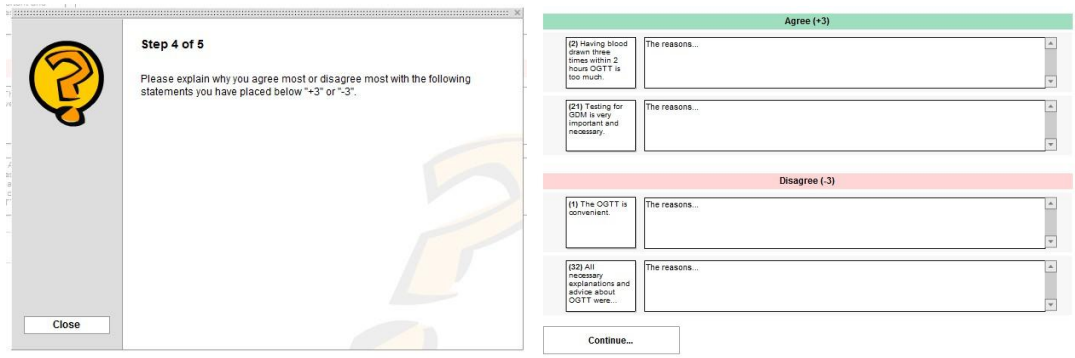


2.2 Rank over the preferences to get a pyramid of (dis)agreements

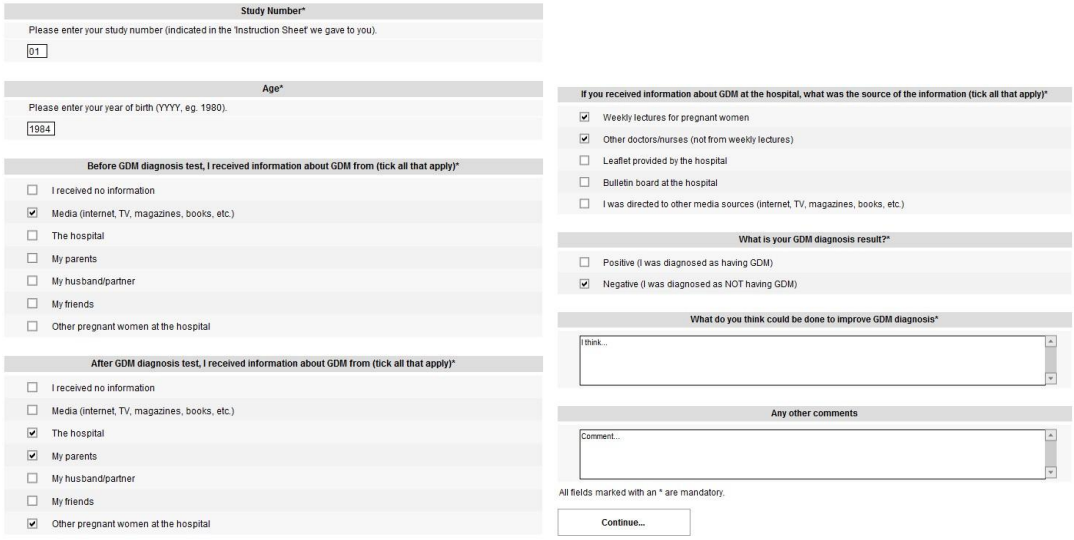




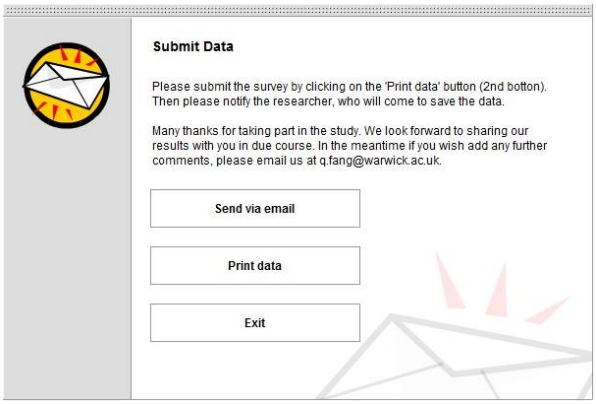
3. Provide reasons for the most agreed and disagreed statements:



4. Provide background information and additional comments:



5. Submit data and finish the FlashQ study:



Appendix 9. An instruction sheet for completing the FlashQ

Thank you very much for participating in the study. Here are some instructions for you before you start!

1) Your study number is: _____ (You will need this information in the last section of the study).
2) The study aims to explore pregnant women's attitudes, views, and experience of the gestational diabetes mellitus (GDM) diagnosis. The study is developed using the FlashQ software.
3) Please use the laptop computer in the meeting room to complete the FlashQ study. The study will take about 15-20 minutes. However, you can take as long as you wish to complete the study.
4) Please simply follow the instructions on the FlashQ to complete. If you have any questions or queries during the process, please do not hesitate to contact the researcher at the other side of the room.
5) When you finish the study, please notify the researcher at the other side of the room. You will receive a small souvenir of the University of Warwick from the researcher.

Thanks again for your precious time!

*If you have any further queries, or if you want to discuss the results of the study, please do not hesitate to contact the researcher. Contact information: Qing Fang (PhD student), Warwick Medical School, the University of Warwick, Coventry CV4 7AL, United Kingdom. Email: Q.Fang@warwick.ac.uk. Contact number: 0086-15810470049 (China); 0044-2476 574505 (United Kingdom).

Appendix 10. Full ethical approval letter for the Q methodology study

7th May 2014

Warwick Medical School

PRIVATE
Ms Qing Fang
176 Cannon Hill Road
Coventry
CV4 7BX

Dear Qing,

Study Title and BSREC Reference: *Pregnant Women's Attitudes, Views, and Experience of the International Association of Diabetes and pregnancy Study Group (IADPSG) Diagnosis Approach for Gestational Diabetes Mellitus in China: a Q Methodology Study*
REGO-2014-704

Thank you for submitting the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Committee for Chair's approval.

I am pleased to confirm that your application meets the required standard which means that full approval is granted and your study may commence.

I take this opportunity to wish you success with the study and to remind you any substantial amendments require approval from the committee before they can be made. Please keep a copy of the original signed version of this letter with your study documentation.

Yours sincerely,

pp

Dr David Davies
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific Research Ethics Subcommittee
A010 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-151875
Email: BSREC@Warwick.ac.uk

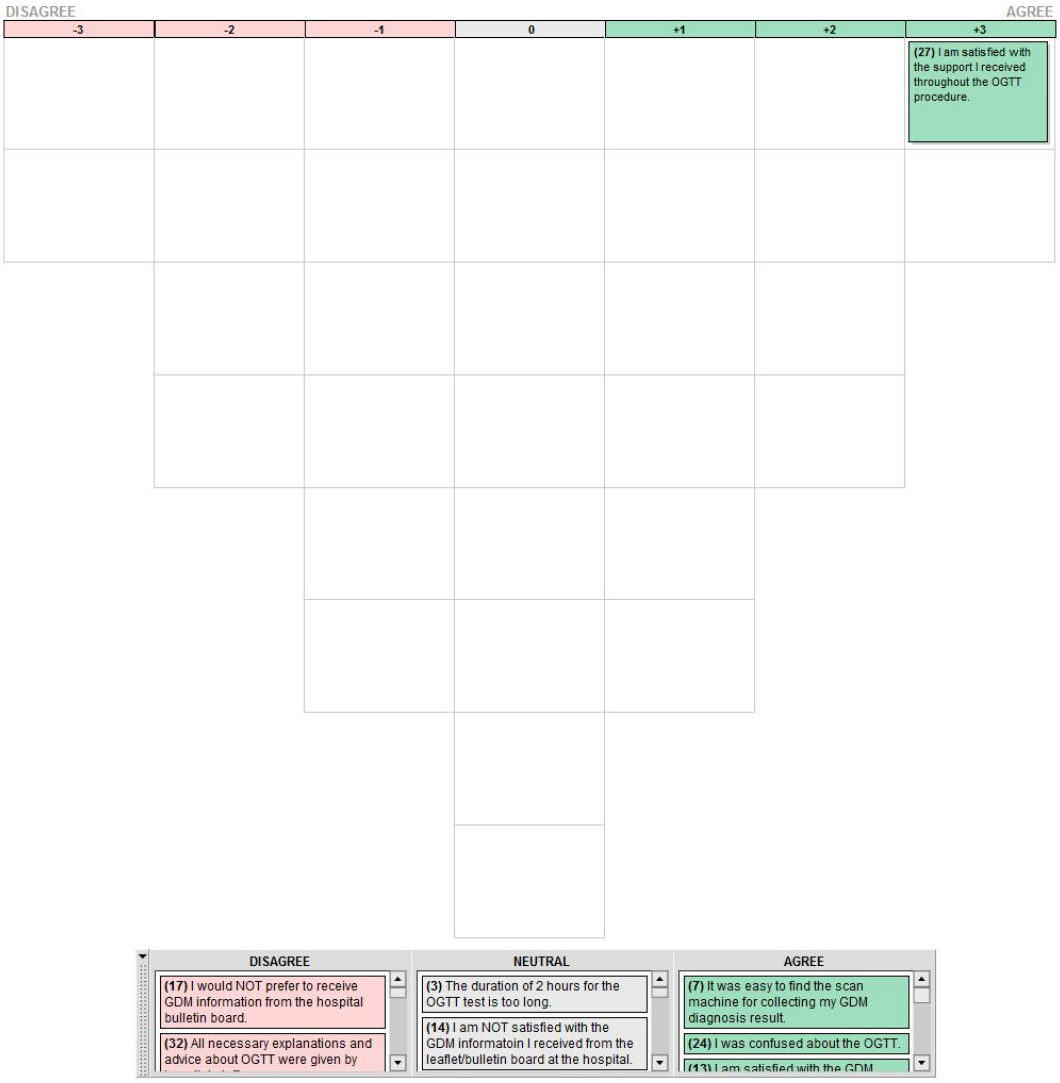
THE UNIVERSITY OF WARWICK

Appendix 11. Table for flagging significant factor loadings (Quick lookup table – is it significant?)

Items	SQRT	1/SQRT	P<0.01	P<0.05
30	5.477225575	0.182574186	0.471	0.358
31	5.567764363	0.179605302	0.463	0.352
32	5.656854249	0.176776695	0.456	0.346
33	5.744562647	0.174077656	0.449	0.341
34	5.830951895	0.171498585	0.442	0.336
35	5.916079783	0.169030851	0.436	0.331
36	6	0.166666667	0.430	0.327
37	6.08276253	0.164398987	0.424	0.322
38	6.164414003	0.162221421	0.419	0.318
39	6.244997998	0.160128154	0.413	0.314
40	6.32455532	0.158113883	0.408	0.310
41	6.403124237	0.156173762	0.403	0.306
42	6.480740698	0.15430335	0.398	0.302
43	6.557438524	0.15249857	0.393	0.299
44	6.633249581	0.150755672	0.389	0.295
45	6.708203932	0.149071198	0.385	0.292
46	6.782329983	0.147441956	0.380	0.289
47	6.8556546	0.145864991	0.376	0.286
48	6.92820323	0.144337567	0.372	0.283
49	7	0.142857143	0.369	0.280
50	7.071067812	0.141421356	0.365	0.277
51	7.141428429	0.140028008	0.361	0.274
52	7.211102551	0.138675049	0.358	0.272
53	7.280109889	0.137360564	0.354	0.269
54	7.348469228	0.136082763	0.351	0.267
55	7.416198487	0.134839972	0.348	0.264
56	7.483314774	0.133630621	0.345	0.262
57	7.549834435	0.132453236	0.342	0.260
58	7.615773106	0.131306433	0.339	0.257
59	7.681145748	0.130188911	0.336	0.255
60	7.745966692	0.129099445	0.333	0.253
61	7.810249676	0.12803688	0.330	0.251
62	7.874007874	0.127000127	0.328	0.249
63	7.937253933	0.125988158	0.325	0.247
64	8	0.125	0.323	0.245
65	8.062257748	0.124034735	0.320	0.243
66	8.124038405	0.123091491	0.318	0.241
67	8.185352772	0.122169444	0.315	0.239
68	8.246211251	0.121267813	0.313	0.238
69	8.306623863	0.120385853	0.311	0.236
70	8.366600265	0.119522861	0.308	0.234
71	8.426149773	0.118678166	0.306	0.233
72	8.485281374	0.11785113	0.304	0.231
73	8.544003745	0.117041147	0.302	0.229
74	8.602325267	0.116247639	0.300	0.228
75	8.660254038	0.115470054	0.298	0.226
76	8.717797887	0.114707867	0.296	0.225

Source: http://jeffar.es/category/q-methodology/

Appendix 12. Screenshot for dragging the Q statements into the Q pyramid



Appendix 13. The list of potential GDM risk factors under investigation in the nested case-control study

Potential GDM risk factors planned to be investigated		Version (electronic, paper, calculated, or N/A)
1	Maternal age (years)	E
2	BMI (kg/m^2) * BMI is calculated as weight (kg)/(height(m))^2	C
3	Family history of diabetes (no family history, father, mother, brother, sister, grandfather, grandmother)	E + P
4	History of GDM (Yes or No)	N/A
5	History of macrosomia (Yes or No)	E + P
6	Unsuccessful pregnancy history (no history, fetal death, stillbirth, abortion, fetal anomaly)	E
7	Gestational weight gain (kg) *Weight gain between pre-pregnancy and OGTT test.	P
8	Polycysticovary syndrome (PCOS) (Yes or No) *PCOS is diagnosed if the patient has all of the following: (1) oligoovulation, (2) signs of androgen excess (clinical or biochemical), (3) other entities are excluded that would cause polycystic ovaries.	P
9	Waist circumference (pre-pregnancy) (cm)	P
10	Height (cm)	E
11	Mother's birth weight (kg)	N/A
12	High blood pressure during first trimester(hypertension, prehypertension, or normal) *Reference value: 120/80 mmHg (normal), 120–139/80–89 mmHg (prehypertension), and140 and/or 90 mmHg or use of antihypertensive medications (hypertension).	P
13	Family history of hypertension (Yes or No)	E + P
14	Irregular menstruation (Yes or No) *Irregular menstruation is a menstrual disorder whose manifestations include irregular cycle lengths as well as metrorrhagia (vaginal bleeding between expected periods).	E
15	Vulvovaginal candidiasis (Yes or No) *Vulvovaginal candidosis is the presence of Candida in addition to vaginal inflammation. The presence of yeast is typically diagnosed in one of three ways: microscopy, microbial culture, and antigen tests.	E + P
16	Triglycerides (above or normal) *Reference value:0.4-1.7mmol/l.	N/A
17	Total cholesterol (above or normal) *Reference value:3.35-6.45mmol/l.	N/A
18	α-thalassaemia (Yes or No) *α -thalassaemiais a form of thalassemia involving the genes HBA1and HBA2. α -thalassaemiais due to impaired production of 1,2,3, or 4 alpha globin chains, leading to a relative excess of beta globin chains.	N/A
19	Hemoglobin (above or normal) *Reference value: 110-150g/L	E + P
20	Elevated serum ferritin level (above or normal) *Reference value: 6.6－28.3 pmol/l.	E + P
21	Maternal Smoking (Yes or No)	E
22	Occupation (unemployed; employed)	E

23	Education level (below college; college or above)	N/A
24	Family income (<¥5000/month; ≥¥5000/month)	N/A
25	Hepatitis B virus carrier (HBsAg+) (Yes or No)	E + P
26	Previous deliveries	E

Appendix 14. Full ethical approval letters for the case-control study

7th May 2014

Warwick Medical School

PRIVATE
Ms Qing Fang
176 Cannon Hill Road
Coventry
CV4 7BX

Dear Qing,

Study Title and BSREC Reference: *A Risk Score System to Improve the International Association of Diabetes and Pregnancy Study groups (IADPSG) Diagnosis Approach for Gestational Diabetes Mellitus in China* **REGO-2014-705**

Thank you for submitting the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Committee for Chair's approval.

I am pleased to confirm that your application meets the required standard which means that full approval is granted and your study may commence.

I take this opportunity to wish you success with the study and to remind you any substantial amendments require approval from the committee before they can be made. Please keep a copy of the original signed version of this letter with your study documentation.

Yours sincerely,

PP

Dr David Davies
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific Research Ethics Subcommittee
A010 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-151875
Email: BSREC@Warwick.ac.uk

THE UNIVERSITY OF WARWICK

Appendix 15. A full list of cut-off scores and corresponding sensitivities and specificities of the ROC curve

Cut-off Score	Sensitivity	1 - Specificity
.0000	1.000	1.000
.1387	1.000	.996
.1615	1.000	.993
.1801	1.000	.989
.1849	.996	.989
.1988	.992	.989
.2087	.989	.989
.2101	.989	.985
.2122	.989	.981
.2164	.985	.981
.2192	.985	.978
.2198	.981	.978
.2251	.981	.974
.2307	.981	.970
.2329	.981	.966
.2351	.981	.963
.2362	.977	.963
.2369	.977	.959
.2377	.977	.955
.2383	.977	.951
.2398	.973	.951
.2423	.973	.948
.2443	.973	.944
.2459	.973	.940
.2484	.973	.936
.2506	.973	.933
.2519	.973	.929
.2527	.970	.929
.2531	.970	.925
.2535	.970	.921
.2541	.970	.918
.2549	.970	.914
.2557	.970	.910
.2560	.966	.910
.2563	.966	.906
.2571	.966	.903
.2587	.966	.899
.2601	.962	.899
.2606	.962	.895

Cut-off Score	Sensitivity	1 - Specificity
.2609	.958	.895
.2640	.958	.891
.2673	.954	.891
.2683	.951	.891
.2695	.947	.891
.2702	.943	.891
.2705	.943	.888
.2708	.939	.888
.2712	.935	.888
.2720	.935	.884
.2733	.935	.880
.2742	.932	.880
.2752	.932	.876
.2763	.932	.873
.2772	.932	.869
.2797	.928	.869
.2828	.928	.865
.2865	.928	.861
.2908	.928	.858
.2932	.924	.858
.2937	.920	.858
.2940	.920	.854
.2943	.920	.850
.2945	.916	.850
.2949	.916	.846
.2955	.916	.843
.2964	.916	.839
.2974	.916	.835
.2982	.916	.831
.2995	.916	.828
.3013	.916	.824
.3042	.916	.820
.3062	.916	.816
.3067	.916	.813
.3078	.916	.805
.3090	.916	.801
.3098	.916	.798
.3113	.916	.794
.3130	.916	.790

Cut-off Score	Sensitivity	1 - Specificity
.3135	.916	.787
.3136	.913	.787
.3144	.909	.787
.3153	.909	.783
.3154	.909	.779
.3154	.909	.775
.3154	.909	.775
.3160	.909	.768
.3170	.905	.768
.3176	.905	.764
.3179	.901	.764
.3188	.901	.760
.3195	.901	.757
.3196	.897	.757
.3209	.897	.753
.3223	.897	.749
.3230	.897	.745
.3238	.894	.745
.3243	.894	.742
.3252	.894	.738
.3263	.890	.738
.3268	.890	.734
.3272	.886	.734
.3275	.886	.730
.3276	.886	.727
.3283	.886	.723
.3290	.886	.719
.3300	.882	.719
.3311	.878	.719
.3315	.878	.715
.3317	.875	.715
.3319	.871	.715
.3324	.871	.712
.3328	.871	.708
.3330	.871	.704
.3333	.867	.704
.3338	.863	.704
.3348	.863	.700

Cut-off Score	Sensitivity	1 - Specificity
.3369	.863	.697
.3392	.863	.693
.3400	.859	.693
.3410	.859	.689
.3428	.859	.685
.3438	.859	.682
.3441	.859	.678
.3457	.859	.674
.3475	.859	.670
.3480	.859	.667
.3483	.859	.663
.3485	.859	.659
.3489	.856	.659
.3493	.856	.655
.3498	.856	.652
.3500	.852	.652
.3505	.852	.644
.3511	.852	.640
.3514	.852	.637
.3524	.852	.633
.3538	.852	.629
.3545	.852	.625
.3549	.848	.625
.3557	.844	.625
.3565	.840	.625
.3567	.840	.622
.3572	.840	.618
.3578	.840	.614
.3582	.840	.610
.3587	.840	.603
.3590	.833	.603
.3593	.833	.599
.3600	.829	.599
.3607	.829	.596
.3609	.829	.592
.3610	.829	.588
.3611	.825	.588
.3614	.821	.588

Cut-off Score	Sensitivity	1 - Specificity
.3624	.821	.584
.3638	.821	.581
.3644	.821	.577
.3647	.817	.577
.3652	.814	.577
.3655	.810	.577
.3659	.810	.573
.3662	.810	.569
.3666	.810	.566
.3669	.810	.562
.3678	.810	.558
.3690	.810	.554
.3701	.810	.551
.3715	.802	.551
.3724	.802	.547
.3729	.798	.547
.3740	.798	.543
.3750	.795	.543
.3760	.795	.539
.3771	.795	.536
.3772	.791	.536
.3781	.787	.536
.3797	.787	.532
.3811	.783	.532
.3819	.783	.528
.3828	.783	.524
.3839	.783	.521
.3841	.779	.521
.3849	.779	.517
.3856	.776	.517
.3858	.768	.517
.3872	.768	.513
.3885	.764	.513
.3886	.764	.509
.3892	.764	.506
.3898	.760	.506
.3910	.760	.502
.3926	.760	.498

Cut-off Score	Sensitivity	1 - Specificity
.3932	.757	.494
.3938	.753	.494
.3941	.749	.494
.3944	.749	.491
.3947	.749	.487
.3948	.749	.483
.3951	.749	.479
.3958	.745	.479
.3969	.745	.476
.3977	.745	.472
.3982	.741	.472
.3989	.738	.472
.4014	.734	.472
.4037	.734	.468
.4047	.734	.464
.4065	.734	.461
.4077	.730	.461
.4079	.730	.457
.4081	.730	.453
.4083	.726	.453
.4090	.726	.449
.4097	.722	.449
.4111	.719	.449
.4125	.715	.449
.4129	.711	.449
.4133	.707	.449
.4139	.707	.446
.4142	.707	.442
.4147	.703	.442
.4156	.700	.442
.4170	.700	.438
.4181	.696	.438
.4184	.696	.434
.4186	.692	.434
.4192	.688	.434
.4199	.684	.434
.4207	.681	.434
.4217	.681	.431

Cut-off Score	Sensitivity	1 - Specificity
.4228	.681	.427
.4244	.681	.423
.4254	.681	.419
.4257	.677	.419
.4259	.677	.416
.4260	.673	.416
.4265	.669	.416
.4274	.665	.416
.4284	.662	.416
.4291	.658	.416
.4294	.658	.412
.4314	.658	.408
.4335	.654	.408
.4336	.650	.408
.4343	.646	.408
.4366	.643	.408
.4387	.643	.404
.4403	.643	.401
.4426	.639	.401
.4440	.639	.397
.4444	.639	.393
.4449	.639	.390
.4453	.639	.386
.4460	.639	.382
.4467	.639	.375
.4471	.639	.371
.4477	.639	.367
.4494	.639	.363
.4510	.635	.363
.4525	.635	.360
.4552	.635	.356
.4570	.635	.352
.4590	.631	.352
.4608	.627	.352
.4610	.624	.352
.4617	.620	.352
.4631	.616	.352
.4642	.612	.352

Cut-off Score	Sensitivity	1 - Specificity
.4649	.612	.348
.4658	.612	.341
.4666	.612	.337
.4668	.612	.333
.4672	.608	.333
.4684	.605	.333
.4703	.601	.333
.4714	.597	.330
.4727	.597	.326
.4746	.597	.322
.4759	.597	.318
.4766	.597	.315
.4778	.597	.311
.4788	.593	.311
.4791	.589	.311
.4796	.589	.307
.4823	.589	.303
.4845	.586	.303
.4849	.582	.303
.4861	.582	.300
.4879	.582	.296
.4903	.578	.296
.4920	.578	.292
.4931	.574	.292
.4941	.574	.288
.4944	.574	.285
.4946	.570	.285
.4949	.570	.281
.4963	.570	.277
.4981	.570	.273
.4989	.567	.273
.4995	.563	.273
.5000	.559	.273
.5009	.559	.270
.5019	.559	.262
.5030	.559	.258
.5042	.559	.255
.5049	.559	.251

Cut-off Score	Sensitivity	1 - Specificity
.5057	.559	.247
.5084	.559	.243
.5105	.555	.243
.5114	.555	.240
.5126	.555	.236
.5131	.551	.236
.5144	.551	.232
.5157	.551	.228
.5175	.548	.228
.5196	.544	.228
.5227	.540	.228
.5255	.536	.228
.5265	.532	.228
.5280	.529	.228
.5291	.525	.228
.5296	.525	.225
.5306	.525	.221
.5317	.521	.221
.5332	.521	.217
.5344	.521	.213
.5346	.517	.213
.5361	.513	.213
.5387	.510	.213
.5400	.506	.210
.5405	.502	.210
.5412	.498	.210
.5419	.498	.206
.5428	.494	.206
.5433	.494	.202
.5438	.490	.202
.5446	.487	.202
.5455	.483	.202
.5482	.479	.202
.5512	.479	.199
.5524	.479	.195
.5544	.475	.195
.5560	.475	.191
.5569	.475	.187

Cut-off Score	Sensitivity	1 - Specificity
.5586	.475	.184
.5607	.471	.184
.5633	.471	.180
.5654	.468	.180
.5664	.464	.180
.5671	.464	.176
.5675	.464	.172
.5691	.464	.169
.5714	.464	.165
.5727	.460	.165
.5733	.460	.161
.5735	.456	.161
.5742	.456	.157
.5765	.452	.157
.5787	.452	.154
.5805	.452	.150
.5833	.452	.146
.5854	.449	.146
.5867	.445	.146
.5895	.441	.146
.5916	.441	.142
.5918	.437	.142
.5936	.437	.139
.5953	.437	.135
.5970	.433	.135
.5987	.433	.131
.6010	.433	.127
.6038	.433	.124
.6056	.430	.124
.6083	.430	.120
.6097	.426	.120
.6104	.426	.116
.6117	.422	.116
.6125	.422	.112
.6130	.418	.112
.6135	.418	.109
.6140	.418	.105
.6143	.414	.105

Cut-off Score	Sensitivity	1 - Specificity
.6148	.411	.105
.6153	.407	.105
.6159	.407	.101
.6172	.403	.101
.6183	.403	.097
.6189	.399	.097
.6196	.395	.097
.6205	.395	.094
.6210	.392	.094
.6228	.388	.094
.6249	.388	.090
.6275	.384	.090
.6298	.380	.090
.6302	.376	.090
.6372	.373	.090
.6446	.369	.090
.6467	.365	.090
.6486	.361	.090
.6499	.357	.090
.6512	.354	.090
.6520	.354	.086
.6527	.350	.086
.6529	.350	.082
.6536	.346	.082
.6551	.346	.079
.6576	.342	.079
.6594	.338	.079
.6597	.335	.079
.6604	.335	.075
.6616	.331	.075
.6623	.331	.071
.6633	.327	.071
.6659	.323	.071
.6680	.319	.071
.6695	.316	.071
.6712	.312	.071
.6729	.308	.071
.6743	.304	.071

Cut-off Score	Sensitivity	1 - Specificity
.6752	.300	.071
.6762	.297	.071
.6774	.297	.067
.6790	.293	.067
.6816	.289	.067
.6837	.285	.067
.6862	.278	.067
.6904	.274	.067
.6923	.270	.067
.6927	.266	.067
.6933	.262	.067
.6939	.259	.067
.6943	.255	.067
.6980	.255	.064
.7017	.255	.060
.7018	.255	.060
.7022	.251	.056
.7051	.247	.056
.7080	.243	.056
.7087	.243	.052
.7091	.240	.052
.7094	.240	.049
.7099	.236	.049
.7115	.232	.049
.7142	.232	.045
.7164	.228	.045
.7186	.224	.045
.7205	.221	.045
.7212	.217	.045
.7222	.213	.045
.7231	.209	.045
.7287	.205	.045
.7341	.202	.045
.7356	.202	.041
.7376	.202	.037
.7432	.198	.037
.7493	.198	.034
.7514	.194	.034

Cut-off Score	Sensitivity	1 - Specificity
.7532	.190	.034
.7573	.190	.030
.7618	.186	.030
.7638	.183	.030
.7643	.179	.030
.7645	.179	.026
.7647	.175	.026
.7660	.171	.026
.7692	.171	.022
.7717	.167	.022
.7723	.163	.022
.7728	.160	.022
.7743	.156	.022
.7757	.152	.022
.7760	.148	.022
.7763	.141	.022
.7780	.141	.019
.7814	.137	.019
.7840	.133	.019
.7857	.129	.019
.7873	.125	.019
.7888	.122	.019
.7903	.118	.019
.7926	.114	.019
.7950	.110	.019
.7974	.106	.019
.8038	.103	.019
.8118	.103	.015
.8157	.099	.015
.8171	.095	.015
.8202	.091	.015
.8222	.087	.015
.8248	.087	.011
.8307	.084	.011
.8349	.080	.011
.8383	.076	.011
.8422	.072	.011
.8442	.068	.011

Cut-off Score	Sensitivity	1 - Specificity
.8482	.065	.011
.8521	.061	.011
.8540	.061	.007
.8568	.061	.004
.8597	.061	.000
.8702	.057	.000
.8814	.053	.000
.8856	.049	.000
.8901	.046	.000
.8928	.042	.000
.8939	.038	.000
.9006	.034	.000
.9084	.030	.000
.9126	.027	.000
.9149	.023	.000
.9171	.019	.000
.9237	.015	.000
.9311	.011	.000
.9426	.008	.000
.9585	.004	.000
1.0000	.000	.000

Appendix 16. Sensitivity and specificity with the two different cut-off scores

3.1 Figures with a cut-off score of 0.32

		IADPSG screening		
		GDM	Non-GDM	Total
Risk scoring algorithm (with a cut-off score of 0.32)	Positive (≥0.32 as having high risk of GDM)	245 (True positive)	211 (False positive)	456
	Negative (<0.32 as having low risk of GDM)	27 (False negative)	67 (True negative)	97
	Total	272	278	553

1. Sensitivity=TP/(TP + FN)=245/272=90%
2. Specificity=TN/(TN + FP)=67/278=24%
3. Sensitivity is the proportion of GDM who are correctly diagnosed out by using the risk scoring algorithm with the IADPSG diagnosis approach.
4. Specificity is the proportion of non-GDM who are classified as low risk women and avoid the IADPSG diagnosis approach.

3.2 Figures with a cut-off score of 0.37

		IADPSG screening		
		GDM	Non-GDM	Total
Risk scoring algorithm (with a cut-off score of 0.37)	Positive (≥0.37 as having high risk of GDM)	218 (True positive)	153 (False positive)	371
	Negative (<0.37 as having low risk of GDM)	54 (False negative)	125 (True negative)	179
	Total	272	278	550

1. Sensitivity=TP/(TP + FN)= 218/272=80%
2. Specificity=TN/(TN + FP)=125/278=45%
3. Sensitivity is the proportion of GDM who are correctly diagnosed out by using the risk scoring algorithm with the IADPSG diagnosis approach.
4. Specificity is the proportion of non-GDM who are classified as low risk women and avoid the IADPSG diagnosis approach.

REFERENCES

Aljohani, N., Rempel, B. M., Ludwig, S., Morris, M., McQuillen, K., Cheang, M., Murray, R. & Shen, G. X. (2008) Gestational diabetes in Manitoba during a twenty-year period. *Clinical and Investigative Medicine*, 31 (3): E131-E137.

Alptekin, H., Çizmecioğlu, A., Işık, H., Cengiz, T., Yildiz, M., & Iyisoy, M. S. (2016) Predicting gestational diabetes mellitus during the first trimester using anthropometric measurements and HOMA-IR. *Journal of endocrinological investigation*, 39(5): 577-583.

Altman, D.G. and Bland, J.M. (1994) Statistics Notes: Diagnostic tests 2: predictive values *British Medical Journal*, 309(6947), p.102.

Anna, V., Van Der Ploeg, H. P., Cheung, N. W., Huxley, R. R. & Bauman, A. E. (2008) Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. *Diabetes care*, 31 (12): 2288-2293.

Association, American Diabetes (1980) American Diabetes Association Workshop-Conference on Gestational Diabetes: summary and recommendations. *Diabetes Care* 3: 499-501.

Association, American Diabetes (1985) Summary and recommendations of the SecondInternational Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes* 34 [Suppl 2]:123-126.

Association, American Diabetes (1997) The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 20:1183-1197.

Association, American Diabetes (2010) Diagnosis and classification of diabetes mellitus. *Diabetes care*, 33 (Supplement 1): S62-S69.

Association, American Diabetes (2013) Diagnosis and classification of diabetes mellitus. *Diabetes care*, 36 (Supplement 1): S67-S74.

Bailey, L., Vardulaki, K., Langham, J. & Chandramohan, D. (2005) *Introduction to epidemiology*. Open University Press.

Baliutaviciene, D., Petrenko, V. & Zalinkevicius, R. (2002) Selective or universal diagnostic testing for gestational diabetes mellitus. *International Journal of Gynaecology & Obstetrics*, 78 (3): 207-211.

Bauman, A.E., Fardy, H.J. and Harris, P.G. (2003) Getting it right: why bother with patient-centred care?. *Medical Journal of Australia*, 179(5), pp.253-256.

Bebbington MW, Milner R, Wilson RD, Harris S. (1999) A randomized controlled trial comparing routine screening vs. selected screening for gestational diabetes in lowrisk population. *Am J Obstet Gynecol*, 180: S36.

Benhalima, K., Hanssens, M., Devlieger, R., Verhaeghe, J. & Mathieu, C. (2013) Analysis of

pregnancy outcomes using the new IADPSG recommendation compared with the carpenter and coustan criteria in an area with a low prevalence of gestational diabetes. *International journal of endocrinology*, 2013.

Brown, S.R. (1980). Political subjectivity: Applications of Q methodology in political sicience. Yale University Press.

Brown, S.R. (1993). A primer on Q methodology. *Operant Subjectivity*, 16 (3/4): 91-138.

Brown, S.R. (1997) The History and Principles of Q methodology in Psychology and the Social Sciences. Department of Political Science, Kent State University, Kent, OH. Available: http://facstaff.uww.edu/cottlec/Qarchive/Bps.htm.

Bryman, A. (2009). Social Research Methods. 3rd ed. Oxford: OXFORD UNIVERSITY PRESS.

Cai, Z. & Yang Z. (2012) The treatment of 112 gestational diabetes mellitus (GDM) cases identified by the IADPSG criteria in China. *Chinese Journal of Birth Health & Heredity*, 8-037. [Original Chinese article: 蔡贞玉&杨隽 (2012) IADPSG 标准诊断妊娠糖尿病患者 (GDM) 112 例临床疗效分析. *中国优生与遗传杂志*, 8-037.]

Caliskan, E., Kayikcioglu, F., Ozturk, N., Koc, S. & Haberal, A. (2004) A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. *Acta Obstetricia et Gynecologica Scandinavica*, 83 (6): 524-530.

Capula, C., Chiefari, E., Vero, A., Arcidiacono, B., Iiritano, S., Puccio, L., Pullano, V., Foti, D. P., Brunetti, A. & Vero, R. (2013) Gestational Diabetes Mellitus: Screening and Outcomes in Southern Italian Pregnant Women. *ISRN endocrinology*, 2013.

Carpenter, M. W. & Coustan, D. R. (1982) Criteria for screening tests for gestational diabetes. *American Journal of Obstetrics & Gynecology*, 144 (7): 768-773.

Casey, B. M., Lucas, M. J., McIntire, D. D. & Leveno, K. J. (1997) Pregnancy outcomes in women with gestational diabetes compared with the general obstetric population. *Obstetrics & Gynecology*, 90 (6): 869-873.

Cattell, Raymond B. (1966) The scree test for the number of factors. Multivariate behavioral research 1.2: 245-276.

Centre for Reviews and Dissemination. (2009). Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: University of York.

Chen, X., Scholl, T. O. & Stein, T. P. (2006) Association of Elevated serum ferritin levels and the risk of gestational diabetes mellitus in pregnant women The Camden Study. Diabetes Care, 29 (5): 1077-1082.

Chinese Ministry of Health (2011). Diagnosis for Gestational Diabetes Mellitus. The Health Standard of People's Republic of China WS 331-2011. [Original Chinese article: 中国卫生部 (2011). 妊娠期糖尿病诊断. 中华人民共和国卫生行业标准 WS 331-2011.]

Chirayath, Haiju Henry. Diabetes management in pregnancy. *Reviews in Gynaecological and*

Perinatal Practice 6.1 (2006): 106-114.

Corcoy, R., Garcia-Patterson, A., Pau, E., Pascual, E., Altirriba, O., Adelantado, J. M. & de Leiva, A. (2004) Is selective screening for gestational diabetes mellitus worthwhile everywhere? *Acta Diabetologica*, 41 (4): 154-157.

Cosson E, Benchimol M, Carbillon L, Paries J, Lormeau B, Sandre-Bacon D, et al (2006). Gestational diabetes mellitus: role of the risk factorson fetal and maternal prognosis. *Diabetes Metab;*32(suppl.1):2611-PO.

Cosson, E., Benbara, A., Pharisien, I., Nguyen, M. T., Revaux, A., Lormeau, B., Sandre-Banon, D., Assad, N., Pillegand, C. & Valensi, P. (2012) Diagnostic and Prognostic Performances Over 9 Years of a Selective Screening Strategy for Gestational Diabetes Mellitus in a Cohort of 18,775 Subjects. *Diabetes Care*.

Cosson, E., Benbara, A., Pharisien, I., Nguyen, M. T., Revaux, A., Lormeau, B., Sandre-Banon, D., Assad, N., Pillegand, C. & Valensi, P. (2013) Diagnostic and Prognostic Performances Over 9 Years of a Selective Screening Strategy for Gestational Diabetes Mellitus in a Cohort of 18,775 Subjects. *Diabetes care*, 36 (3): 598-603.

Crowther, C. A., Hiller, J. E., Moss, J. R., McPhee, A. J., Jeffries, W. S. & Robinson, J. S. (2005) Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. *New England Journal of Medicine*, 352 (24): 2477-2486.

Cross, R.M. (2005). Exploring attitudes: The case for Q methodology. *Health Education Research.* 20(2): 206-213.

Cundy, T., Ackermann, E. & Ryan, E. A. (2014) Gestational diabetes: new criteria may triple the prevalence but effect on outcomes is unclear. *BMJ: British Medical Journal*, 348.

Damm, P., Hod, M., Lefebvre, P. & Wolfson, E. (2009) Future risk of diabetes in mother and child after gestational diabetes mellitus. Conference Proceedings.

Danilenko-Dixon, D. R., Van Winter, J. T., Nelson, R. L. & Ogburn Jr, P. L. (1999) Universal versus selective gestational diabetes screening: Application of 1997 American Diabetes Association recommendations. *American Journal of Obstetrics and Gynecology*, 181 (4): 798-802.

Davey, R. X. & Hamblin, P. S. (2001) Selective versus universal screening for gestational diabetes mellitus: an evaluation of predictive risk factors. *Medical journal of Australia*, 174 (3): 118-121.

Di Cianni G, Miccoli R, Volpe L, Lencioni C, Del Prato S: Intermediate metabolism in normal pregnancy and in gestational diabetes. *Diabetes Metab Res Rev* 19:259-270, 2003.

Di Cianni, G., Volpe, L., Lencioni, C., Miccoli, R., Cuccuru, I., Ghio, A., Chatzianagnostou, K., Bottone, P., Teti, G. & Del Prato, S. (2003) Prevalence and risk factors for gestational diabetes assessed by universal screening. *Diabetes research and clinical practice*, 62 (2): 131.

Dietrich, M. L., Dolnicek, T. F. & Rayburn, W. F. (1987) Gestational diabetes screening in a private, midwestern American population. *American Journal of Obstetrics and Gynecology*, 156

(6): 1403-1408.

Dornhorst, A., Paterson, C., Nicholls, J., Wadsworth, J., Chiu, D., Elkeles, R., Johnston, D. & Beard, R. (1992) High prevalence of gestational diabetes in women from ethnic minority groups. *Diabetic Medicine*, 9 (9): 820-825.

Downs, S. H. & Black, N. (1998) The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of epidemiology and community health*, 52 (6): 377-384.

Drummond (1996) Guidelines for authors and peer reviewers of economic submissions to the BMJ. *British Medical Journal*; 313: 275.

Du, Y. (2013) Clinical analysis of the risk factors for gestational diabetes mellitus. Inner *Mongol Journal of Traditional Chinese Medicine*, 2013 (13): 29-20. [Original Chinese article: 杜玉华 (2013) 妊娠糖尿病危险因素的临床分析. *内蒙古中医药*, 2013 (13): 29-20.]

Ernster, V. L. (1994) Nested case-control studies. *Preventive medicine*, 23 (5): 587-590.

Ezimokhai, M., Joseph, A. & Bradley-Watson, P. (2006) Audit of pregnancies complicated by diabetes from one center five years apart with selective versus universal screening. *Annals of the New York Academy of Sciences*, 1084 132-140.

Falavigna, M., Schmidt, M. I., Trujillo, J., Alves, L. F., Wendland, E. R., Torloni, M. R., Colagiuri, S. & Duncan, B. B. (2012) Effectiveness of gestational diabetes treatment: a systematic review with quality of evidence assessment. *Diabetes Research and Clinical Practice*, 98 (3): 396-405.

Ferrara, A. (2007) Increasing prevalence of gestational diabetes mellitus a public health perspective. *Diabetes Care*, 30 (Supplement 2): S141-S146.

Force, U. S. P. S. T. (2008) Screening for gestational diabetes mellitus: U.S. Preventive Services Task Force recommendation statement.[Summary for patients in Ann Intern Med. 2008 May 20;148(10):I60; PMID: 18490671]. *Annals of Internal Medicine*, 148 (10): 759-765.

Griffin, M. E., Coffey, M., Johnson, H., Scanlon, P., Foley, M., Stronge, J., O'Meara, N. M. & Firth, R. G. (2000) Universal vs. risk factor-based screening for gestational diabetes mellitus: detection rates, gestation at diagnosis and outcome. *Diabetic Medicine*, 17 (1): 26-32.

Griffiths, R. D., Rodgers, D. V. & Moses, R. G. (1993) Patients' attitudes toward screening for gestational diabetes mellitus in the Illawarra area, Australia. *Diabetes Care*, 16 (2): 506-508.

Group, N. D. D. (1979) Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes*, 28 (12): 1039-1057.

Guttman, Louis. Some necessary conditions for common-factor analysis. *Psychometrika* 19.2 (1954): 149-161.

Gwet, K. (2010) Handbook of Inter-Rater Reliability: The Definite Guide to Measuring the Extent of Agreement among Raters.

Hanley, J. A. & McNeil, B. J. (1983) A method of comparing the areas under receiver operating

characteristic curves derived from the same cases. *Radiology*, 148 (3): 839-843.

Hartling, L., *et al.* (2012) Screening and Diagnosing Gestational Diabetes Mellitus. *Evidence Reports/Technology Assessments*, No. 210. Available from: http://www.ncbi.nlm.nih.gov/books/NBK114844/pdf/TOC.pdf

Harre, F. E., Lee, K. L. & Pollock, B. G. (1988) Regression models in clinical studies: determining relationships between predictors and response. *Journal of the National Cancer Institute*, 80 (15): 1198-1202.

Hieronimus, S. & Le Meaux, J. P. (2010) Relevance of gestational diabetes mellitus screening and comparison of selective with universal strategies. *Diabetes & Metabolism*, 36 (6 Pt 2): 575-586.

Holt, R. I. G., *et al.* The effectiveness of glibenclamide in women with gestational diabetes. *Diabetes, Obesity and Metabolism* 10.10 (2008): 906-911.

Hou, M., Wang, Z., Zhou, L., Shi, L., & Qiao, Q. (2012) The risk factors for gestional diabetes mellitus under the new IADPSG diagnostic criteria. *Progress in Modern Biomedicine ISTIC*, 12 (10). [Original Chinese article: 侯美芹, 王治洁, 周玲, 石礼红&乔侨 (2012) 新诊断标准下妊娠期糖尿病高危因素研究. *现代生物医学进展*, ISTIC, 12 (10).]

Jensen, D. M., Molsted-Pedersen, L., Beck-Nielsen, H., Westergaard, J. G., Ovesen, P. & Damm, P. (2003) Screening for gestational diabetes mellitus by a model based on risk indicators: A prospective study. *American Journal of Obstetrics and Gynecology*, 189 (5): 1383-1388.

Jiang, Y., You, Y., & Zhang, L. (2013) The clinical value of the new IADPSG diagnostic criteria for gestational diabetes mellitus. *National Medical Frontiers of China*, 8 (1): 6-7. [Original Chinese article: 蒋玉蓉, 游一平&张丽娟 (2013) IADPSG 糖尿病诊断新标准的临床应用价值. *中国医疗前沿*, 8 (1): 6-7.]

Jiao, Z. (2003) The risk factors for gestational diabetes mellitues and the molecular epidemiology study on GCK and CTLA4 genes. [Original Chinese article: 焦振山 (2003) 妊娠糖尿病危险因素及 GCK 和 CTLA4 基因的分子流行病学研究.]

Jimenez-Moleon, J. J., Bueno-Cavanillas, A., Luna-del-Castillo, J. D., Garcia-Martin, M., Lardelli-Claret, P. & Galvez-Vargas, R. (2002) Prevalence of gestational diabetes mellitus: Variations related to screening strategy used. *European Journal of Endocrinology*, 146 (6): 831-837.

Jiwani, A., Marseille, E., Lohse, N., Damm, P., Hod, M. & Kahn, J. G. (2012) Gestational diabetes mellitus: Results from a survey of country prevalence and practices. *Journal of Maternal-Fetal and Neonatal Medicine*, 25 (6): 600-610.

Ju, H., Rumbold, A. R., Willson, K. J. & Crowther, C. A. (2008) Borderline gestational diabetes mellitus and pregnancy outcomes. *BMC Pregnancy and Childbirth*, 8 (1): 31.

Kaiser, Henry F. (1960) The application of electronic computers to factor analysis. *Educational*

and psychological measurement.

Kaiser, Henry F. (1970) A second generation little jiffy. *Psychometrika* 35.4: 401-415.

Khan, K. S., Ter Riet, G., Glanville, J., Sowden, A. J. & Kleijnen, J. (2001) *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews.* NHS Centre for Reviews and Dissemination.

Kim, C., Newton, K. M. & Knopp, R. H. (2002) Gestational Diabetes and the Incidence of Type 2 Diabetes A systematic review. *Diabetes care*, 25 (10): 1862-1868.

Kim C., Newton K.M., Knopp R.H. (2002) Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care*, 25:1862-1868.

Landon, M. B., Spong, C. Y., Thom, E., Carpenter, M. W., Ramin, S. M., Casey, B., Wapner, R. J., Varner, M. W., Rouse, D. J. & Thorp Jr, J. M. (2009) A multicenter, randomized trial of treatment for mild gestational diabetes. *New England Journal of Medicine*, 361 (14): 1339-1348.

Langer, O., Yogev, Y., Most, O. & Xenakis, E. M. (2005) Gestational diabetes: the consequences of not treating. *American journal of obstetrics and gynecology*, 192 (4): 989-997.

Langer, O., Umans, J. G. & Miodovnik, M. (2013) Perspectives on the proposed gestational diabetes mellitus diagnostic criteria. *Obstetrics and Gynecology*, 121 (1): 177-182.

Lao, T. & Ho, L. (2001) α-Thalassaemia trait and gestational diabetes mellitus in Hong Kong. *Diabetologia*, 44 (8): 966-971.

Lao, T., Chan, P. & Tam, K. (2001) Gestational diabetes mellitus in the last trimester-a feature of maternal iron excess? *Diabetic medicine*, 18 (3): 218-223.

Lao, T. T., Chan, L. Y., Tam, K.-F. & Ho, L.-F. (2002) Maternal hemoglobin and risk of gestational diabetes mellitus in Chinese women. *Obstetrics & Gynecology*, 99 (5, Part 1): 807-812.

Lao, T. T. & Ho, L.-F. (2003) First-trimester blood pressure and gestational diabetes in high-risk Chinese women. *Journal of the Society for Gynecologic Investigation*, 10 (2): 94-98.

Lao, T. T., Chan, B. C., Leung, W.-C., Ho, L.-F. & Tse, K.-Y. (2007) Maternal hepatitis B infection and gestational diabetes mellitus. *Journal of hepatology*, 47 (1): 46-50.

Lapolla, A., Dalfrà, M., Ragazzi, E., De Cata, A. & Fedele, D. (2011) New International Association of the Diabetes and Pregnancy Study Groups (IADPSG) recommendations for diagnosing gestational diabetes compared with former criteria: a retrospective study on pregnancy outcome. *Diabetic Medicine*, 28 (9): 1074-1077.

Lavin, J. P., Barden, T. P., and Miodovnik, M. (1981) Clinical experience witha screening program for gestational diabetes. *Am. J. Obstet. Gynecol*, 141:491-94.

Le, J. (2008) Gynecology and Obstetrics (7th edition). Beijing: People's medical publishing house. 150 -154. [Original Chinese article: 乐杰 (2008) 妇产科学(7 版). 北京: 人民卫生出版社. 150 -154.]

Lowe Jr, W. L., Scholtens, D. M., Sandler, V., & Hayes, M. G. (2016) Genetics of gestational

diabetes mellitus and maternal metabolism. *Current diabetes reports*, 16(2): 1-10.

Luesley, David M., and Mark D. Kilby, eds. *Obstetrics & Gynaecology*: An Evidence-based Text for MRCOG. CRC Press, 2016.

Lu, Y., Wang, J., Lu, L., Chen, W., & Li, Y. (2012) Comments and exploration of the new IADPSG diagnostic critieria. *Chinese Journal of Laboratory Diagnosis*, 16 (6): 1041-1043. [Original Chinese article: 芦雅苹, 王加, 陆琳琳, 陈伟萍&李艳芳 (2012) IADPSG 诊断标准用于妊娠期糖尿病诊断的探讨. *中国实验诊断学*, 16 (6): 1041-1043.]

Ma, R. M., Lao, T. T., Ma, C. L., Liao, S. J., Lu, Y. F., Du, M. Y., Xiao, H., Zhang, L., Yang, M. H. & Xiao, X. (2007) Relationship between leg length and gestational diabetes mellitus in Chinese pregnant women. *Diabetes care*, 30 (11): 2960-2961.

McKeown, B. & Thomas, D. (1988) *Q methodology*. Sage.

Meek, C.L., Murphy, H.R. and Simmons, D. (2016) Random plasma glucose in early pregnancy is a better predictor of gestational diabetes diagnosis than maternal obesity. *Diabetologia*, 59(3), pp.445-452.

Metzger BE (1991) Summary and recommendations of theThird International Workshop-Conference on Gestational Diabetes Mellitus. *Diabetes* 40 [Suppl 2]: 197-201.

Metzger, B. E. & Coustan, D. R. (1998) Summary and recommendations of the fourth international workshop- conference on gestational diabetes mellitus. *Diabetes Care*, 21 (SUPPL.2): B161-B167.

Metzger, B. E., Buchanan, T. A., Coustan, D. R., De Leiva, A., Dunger, D. B., Hadden, D. R., Hod, M., Kitzmiller, J. L., Kjos, S. L. & Oats, J. N. (2007) Summary and recommendations of the fifth international workshop-conference on gestational diabetes mellitus. *Diabetes care*, 30 (Supplement 2): S251-S260.

Minsart, A. F., Lescrainier, J. P. & Vokaer, A. (2009) Selective versus Universal Screening for Gestational Diabetes Mellitus: An Evaluation of Naylor's Model. *Gynecologic and Obstetric Investigation*, 68 (3): 154-159.

Mission, J. F., Ohno, M. S., Cheng, Y. W. & Caughey, A. B. (2012) Gestational diabetes screening with the new IADPSG guidelines: A cost-effectiveness analysis. *American Journal of Obstetrics and Gynecology*, 207 (4): 326.e321-326.e329.

Mohamed Ismail, N. A., Mohd Kasim, M., Noor Aizuddin, A., & Umar, N. A. (2013) Homeostatic indices of insulin resistance among gestational diabetics in anticipating pregnancy complications. *Gynecological Endocrinology*, 29(7): 691-694.

Moher, D. et al (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRlSMA Statement. PLoS Med 6(6): e I000097.

Moses, R. G., Moses, J. & Davis, W. S. (1998) Gestational diabetes: Do lean young caucasian women need to be tested? *Diabetes Care*, 21 (11): 1803-1806.

Moyer, V. A. (2014) Screening for Gestational Diabetes Mellitus: US Preventive Services Task Force Recommendation Statement. *Annals of internal medicine.*

National Institutes of Health. (2013) NIH Consensus Development Conference on Diagnosing Gestational Diabetes Mellitus. NIH Consensus Development Conference Statements, 29 (Number 1) March 4-6, 2013 [Cited March 2014]. Available: http://prevention.nih.gov/cdp/conferences/2013/gdm/files/Gestational_Diabetes_Mellitus508.pdf

Naylor, C. D., Sermer, M., Chen, E. & Farine, D. (1997) Selective screening for gestational diabetes mellitus. *New England Journal of Medicine,* 337 (22): 1591-1596.

NICE (2008) Diabetes in Pregnancy: Management of Diabetes and Its Complications from Preconception to the Postnatal Period.NICE Clinical Guidelines, No. 63.*National Collaborating Centre for Women's and Children's Health (UK).*London: RCOG Press.

NICE (2015) NICE guidance on diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. NICE clinical guideline NG3. London, February 2015. Available from: nice.org.uk/guidance/ng3.

Nielsen, K. K., de Courten, M. & Kapur, A. (2012) The urgent need for universally applicable simple screening procedures and diagnostic criteria for gestational diabetes mellitus--lessons from projects funded by the World Diabetes Foundation. *Global health action,* 5.

O'SULLIVAN, J. B. & Mahan, C. M. (1964) Criteria for the oral glucose tolerance test in pregnancy. *Diabetes,* 13 278.

O'Sullivan JB, Mehan CM, Charles D, Dandrow RV (1973) Screening criteria for high-risk gestational diabetic patents. *Am J Obstet Gynecol*, 116: 895-900.

Odar, E., Wandabwa, J. & Kiondo, P. (2004) Maternal and fetal outcome of gestational diabetes mellitus in Mulago Hospital, Uganda. *African Health Sciences,* 4 (1): 9-14.

Organization, W. H. (1999) Definition, diagnosis and classification of diabetes mellitus and its complications.Report of a WHO consultation. Part 1: diagnosis and classification of diabetes mellitus. Geneva:World Health Organization.

Ostlund, I. & Hanson, U. (2003) Occurrence of gestational diabetes mellitus and the value of different screening indicators for the oral glucose tolerance test. *Acta Obstetricia Et Gynecologica Scandinavica*, 82 (2): 103-108.

Panel, I. C. (2010) International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*, 33 (3): 676-682.

Perkins, J. M., Dunn, J. P. & Jagasia, S. M. (2007) Perspectives in gestational diabetes mellitus: A review of screening, diagnosis, and treatment. *Clinical Diabetes,* 25 (2): 57-62.

Phaloprakarn, C., Tangjitgamol, S. & Manusirivithaya, S. (2009) A risk score for selective screening for gestational diabetes mellitus. *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 49 (5): 572-572.

Phaloprakarn, C., Tangjitgamol, S. & Manusirivithaya, S. (2009) A risk score for selective

screening for gestational diabetes mellitus. *European Journal of Obstetrics Gynecology and Reproductive Biology*, 145 (1): 71-75.

Poncet, B., Touzet, S., Rocher, L., Berland, M., Orgiazzi, J. & Colin, C. (2002) Cost-effectiveness analysis of gestational diabetes mellitus screening in France. *European Journal of Obstetrics, Gynecology, & Reproductive Biology*, 103 (2): 122-129.

Poyhonen-Alho, M. K., Teramo, K. A., Kaaja, R. J. & Hiilesmaa, V. K. (2005) 50gram oral glucose challenge test combined with risk factor-based screening for gestational diabetes. *European journal of obstetrics, gynecology, and reproductive biology*, 121 (1): 34-37.

PQMethod Manual (version 2.35, March 2014). Available: http://schmolck.org/qmethod/pqmanual.htm.

Prevention, I. & TYPE, D. O. (2011) Standards of Medical Care in Diabetes-2011. *Diabetes Care*, 34 S11.

Prentice, R. L. (1986) A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73 (1): 1-11.

Qian, L. (2012) A case-control study of the risk factors for gestational diabetes mellitues. *Tianjin Medical Journal*, 40 (9): 951-952. [Original Chinese article: 钱丽雅 (2012) 妊娠糖尿病危险因素的病例对照研究. *天津医药*, 40 (9): 951-952.]

Qiang, Z., Wang, J., & Qi, X. (2006) A case-control study of the risk factors for gestational diabetes mellitues in China. *Chinese Journal Of Public Health*, 22 (7): 795-796. [Original Chinese article: 强兆艳, 王建华&齐秀英 (2006) 妊娠糖尿病危险因素病例对照研究. *中国公共卫生*, 22 (7): 795-796.]

Raspe, Heiner. Prioritisation: (At Least) Two Normative Cultures. *Prioritization in Medicine*. Springer International Publishing, 2016. 85-99.

Ratner, R.E., Christophi, C.A., Metzger, B.E., Dabelea, D., Bennett, P.H., Pi-Sunyer, X., Fowler, S. and Kahn, S.E. (2008) Prevention of diabetes in women with a history of gestational diabetes: effects of metformin and lifestyle interventions. The *Journal of Clinical Endocrinology & Metabolism*, 93(12), pp.4774-4779.

Reece, E. A., Leguizamon, G. & Wiznitzer, A. (2009) Gestational diabetes: the need for a common ground. *The Lancet*, 373 (9677): 1789-1797.

Ren, X., Huang, S., Shao, J., Yan, W., & Zhou, P. (2008) Analysis of the incidence and risk factors for gestational diabetes mellitues. *Maternal and Child Health Care of China*, 23 (21): 2954-2956. [Original Chinese article: 任香梅, 黄水平, 邵继红, 严文君&周萍 (2008) 妊娠糖尿病发病率及危险因素分析. *中国妇幼保健*, 23 (21): 2954-2956.]

Rowan, J. A., Hague, W. M., Gao, W., Battin, M. R. & Moore, M. P. (2008) Metformin versus insulin for the treatment of gestational diabetes. *New England Journal of Medicine*, 358 (19): 2003-2015.

Russell, M. A., Carpenter, M. W. & Coustan, D. R. (2007) Screening and diagnosis of gestational diabetes mellitus. *Clinical Obstetrics and Gynecology*, 50 (4): 949-958.

Sachse, D., Sletner, L., Mørkrid, K., Jenum, A. K., Birkeland, K. I., Rise, F., ... & Berg, J. P. (2012) Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes. *PloS one*, 7(12): e52399.

Sacks, D. A., Abu-Fadil, S., Karten, G. J., Forsythe, A. B. & Hackett, J. R. (1987) Screening for gestational diabetes with the one-hour 50-g glucose test. *Obstetrics & Gynecology*, 70 (1): 89-93.

Scott, D. A., Loveman, E., McIntyre, L. & Waugh, N. (2002) Screening for gestational diabetes: a systematic review and economic evaluation. *Health Technology Assessment (Winchester, England)*, 6 (11): 1-161.

Shang, M. & Ma, L. (2010) The adoption of the IADPSG diagnostic criteria for gestational diabetes mellitues in Beijing. *Chinese Journal of Clinical Obstetrics and Gynecology*, 12 (3): 168-171. [Original Chinese article: 商敏 & 马丽 (2011) IADPSG 诊断标准用于北京市妊娠期糖尿病诊断的探讨. *中国妇产科临床杂志*, 12 (3): 168-171.]

Simmons, D., Devers, M. C., Wolmarans, L. & Johnson, E. (2009) Difficulties in the use of risk factors to screen for gestational diabetes mellitus. *Diabetes Care*, 32 (1): e8.

Smith, N.W. (2001). Current systems in psychology: history, theory, research, and applications. Wadsworth.

Stainton Rogers, R. (1995) Q methodology. *Rethinking methods in psychology*, 178-192.

Stephenson, W. (1953) The study of behavior; Q-technique and its methodology. Chicago: University of Chicago Press.

Stiggelbout, A.M., Van der Weijden, T., De Wit, M.P., Frosch, D., Légaré, F., Montori, V.M., Trevena, L. and Elwyn, G. (2012) Shared decision making: really putting patients at the centre of healthcare. *British Medical Journal*, 344(S 28).

Summary and Recommendations of the2nd International Workshop-Conferenceon Gestational Diabetes Mellitus. Diabetes34 (Suppl. 2):123-126, 1985.

Swinker, M. (1983) Routine screening for gestational diabetes mellitus in a family practice center. *The Journal of family practice*, 17 (4): 611.

Teh, W. T., Teede, H. J., Paul, E., Harrison, C. L., Wallace, E. M. & Allan, C. (2011) Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. *Australian & New Zealand Journal of Obstetrics & Gynaecology*, 51 (1): 26-30.

Test, O. G.-T. (2008) Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*, 2008 (358): 1991-2002.

The National Health Service (2012). Gestational diabetes [cited Dec 11, 2012]. Available: http://www.nhs.uk/Conditions/gestational-diabetes/Pages/Introduction.aspx.

Thomas DB & Baas LR. The issue of generelization in Q methodology: "reliable schematics"

revisited. *Operant Subjectivity* 1992;16(1): 18-36.

Tieu, J., Middleton, P., McPhee, A. J. & Crowther, C. A. (2010) Screening and subsequent management for gestational diabetes for improving maternal and infant health. *Cochrane Database of Systematic Reviews*, (7).

Tsatsoulis, A., Wyckoff, J. & Brown, F. M. (2009) *Diabetes in Women: Pathophysiology and Therapy*. Springer.

Turok, D. K., Ratcliffe, S. D. & Baxley, E. G. (2003) Management of Gestational Diabetes Mellitus. *American Family Physician*, 68 (9): 1767-1772+1775-1776.

U.S. Preventive Services Task Force. (2013) Screening for Gestational Diabetes Mellitus: U.S. Preventive Services Task Force Recommendation Statement. [cited Aug 15, 2013]. Available from: http://www.uspreventiveservicestaskforce.org/uspstf13/ gdm/ gdmdraftrec.htm

Vambergue, A. (2010) Expert consensus on gestational diabetes mellitus. *Diabetes &Metabolism*, 36 (6 Pt 2): 511.

Vandorsten, J., Dodson, W., Espeland, M., Grobman, W., Guise, J., Mercer, B., Minkoff, H., Poindexter, B., Prosser, L. & Sawaya, G. (2012) NIH Consensus Development Conference: Diagnosing Gestational Diabetes Mellitus. *NIH consensus and state-of-the-science statements*, 29 (1): 1-31.

Van Exel, J. & de Graaf, G. (2005) Q methodology: A sneak preview. *Online document: http://www. qmethodology. net/PDF/Q-methodology*.

Van Leeuwen, M., Opmeer, B. C., Zweers, E. J. K., Van Ballegooie, E., Ter Brugge, H. G., De Valk, H. W., Visser, G. H. A. & Mol, B. W. J. (2010) Estimating the risk of gestational diabetes mellitus: A clinical prediction model based on patient characteristics and medical history. *BJOG: An International Journal of Obstetrics and Gynaecology*, 117 (1): 69-75.

Vidaeff, A. C., Yeomans, E. R. & Ramin, S. M. (2003) Gestational diabetes: a field of controversy. *Obstetrical & Gynecological Survey*, 58 (11): 759-769.

Wacholder, S. (1991) Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology*, 2 (2): 155-158.

Walker, J. (2008) NICE guidance on diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period. NICE clinical guideline 63. London, March 2008. *Diabetic Medicine*, 25 (9): 1025-1027.

Ward, W. (2010) Q and you: The application of Q methodology in recreation research.

Watts, S. & Stenner, P. (2005) Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology*, 2 (1): 67-91.

Watts, Simon, and Paul Stenner. (2012) *Doing Q methodological research: Theory, method & interpretation*. Sage.

Waugh, N., Pearson, D. & Royle, P. (2010) Screening for hyperglycaemia in pregnancy: Consensus and controversy. *Best Practice & Research Clinical Endocrinology & Metabolism*,

24 (4): 553-571.

Waugh, N., Royle, P., Clar, C., Henderson, R., Cummins, E., Hadden, D. R., Lindsay, R. & Pearson, D. (2010) Screening for hyperglycaemia in pregnancy: a rapid update for the National Screening Committee. *Health technology assessment*, 14 (45): 1-183.

Webber, Jonathan, Mary Charlton, and Nina Johns. Diabetes in pregnancy: management of diabetes and its complications from preconception to the postnatal period (NG3). *British Journal of Diabetes* 15.3 (2015): 107-111.

Wei, Y. & Yang, H. (2011) Comparison of different diagnostic critiera for gestational diabetes mellitues. *Chinese Journal of Obstetrics and Gynecology*, 46 (8): 578-581. [Original Chinese article: 魏玉梅&杨慧霞 (2011) 妊娠期糖尿病不同诊断标准适宜性的 比较. *中华妇产科杂志*, 46 (8): 578-581.]

Werner E, Pettker C, Zuckerwise L, Reel M, Funai E. (2012) Screening for gestational diabetes mellitus: are the criteria proposed by the International Association of the Diabetes and Pregnancy Study Groups cost-effective? *Diabetes Care* (35):529-35.

WHO (2013). WHO diagnostic criteria - Diabetes Research andClinical Practice [EB/OL].[2013-08-01].[cited March 2014]. Available: http://www.who.int/diabetes/publications/Hyperglycaemia_In_Pregnancy/en/index.html

Williams, C. B., Iqbal, S., Zawacki, C. M., Yu, D. H., Brown, M. B. & Herman, W. H. (1999) Effect of selective screening for gestational diabetes. *Diabetes Care*, 22 (3): 418-421.

Wilson, I.B. (2005). Person-place engagement among recreation visitors: A Q-method inquiry. (Doctoral dissertation, Oklahoma State University). *Dissertation Abstracts International*, B 66/02, 788.

Williams, C. B., Iqbal, S., Zawacki, C. M., Yu, D., Brown, M. B. & Herman, W. H. (1999) Effect of selective screening for gestational diabetes. *Diabetes care*, 22 (3): 418-421.

Wong, Vincent W. Diagnosis of Gestational Diabetes Mellitus: Where are we at?. *International Journal of Diabetes & Clinical Diagnosis* 2014 (2014).

Xiang, A. H., Peters, R. K., Kjos, S. L., Marroquin, A., Goico, J., Ochoa, C., Kawakubo, M. & Buchanan, T. A. (2006) Effect of pioglitazone on pancreatic β-cell function and diabetes risk in Hispanic women with prior gestational diabetes. *Diabetes*, 55 (2): 517-522.

Yang, H., Wei, Y., Gao, X., Xu, X., Fan, L., He, J., Hu, Y., Liu, X., Chen, X., Yang, Z. & Zhang, C. (2009) Risk factors for gestational diabetes mellitus in Chinese women - A prospective study of 16 286 pregnant women in China. *Diabetic Medicine*, 26 (11): 1099-1104.

Yang, H., Zhang, M., Sun, W. & Zhao, Y. (2005) Factors for the impaired glucose tolerance during pregancy. *Chinese Journal of Obstetrics and Gynecology*, 40 (11): 725-728. [Original Chinese article: 杨慧霞, 张眉花, 孙伟杰&赵怿 (2005) 妊娠期糖代谢异常相关因素的研究. *中华妇产科杂志*, 40 (11): 725-728.]

Yang, H. (2009) Recommendation guideline on the diagnosis and treatment of gestational

diabetes mellitues (draft). *Drug Evaluation*, (8): 310-312. [Original Chinese article: 杨慧霞 (2009) 妊娠合并糖尿病临床诊断与治疗推荐指南 (草案). *药品评价*, (8): 310-312.]

Yang, X., Hsu-Hage, B., Zuang, H., Yu, L., Dong, L., Li, J., Shao, P. & Zhang, C. (2002) Gestational diabetes mellitus in women of single gravidity in Tianjin City, China. *Diabetes Care*, 25 (5): 847-851.

Yang, X., Hsu-Hage, B., Yu, L. & Simmons, D. (2002) Selective screening for gestational diabetes in Chinese women. *Diabetes care*, 25 (4): 796.

Yuan, L. (2008) Analysis of the risk factors for gestational diabetes mellitues. *Morden Hospital*, 8 (6): 41-42. [Original Chinese article: 袁丽芳 (2008) 妊娠期糖尿病危险因素的临床分析. *现代医院*, 8 (6): 41-42.]